UNIVERSITÄTSKLINIKUM HAMBURG-EPPENDORF

Klinik und Poliklinik für Psychosomatische Medizin und Psychotherapie

Prof. Dr. med. Dipl.-Psych. Bernd Löwe

Benefits and harms of internet-based depression screening followed by automated feedback of screening results: A mixed-methods interventional evaluation in adults with suspected but undiagnosed depressive disorder

Dissertation

zur Erlangung des Doktorgrades Dr. rer. biol. hum.
an der Medizinischen Fakultät der Universität Hamburg.

vorgelegt von:

Franziska Sikorski
aus Hannover

Hamburg 2025

Angenommen von der
Medizinischen Fakultät der Universität Hamburg am: 04. Juli 2025

Veröffentlicht mit Genehmigung der
Medizinischen Fakultät der Universität Hamburg.

Prüfungsausschuss, der/die Vorsitzende: Prof. Dr. Sebastian Kohlmann

Prüfungsausschuss, zweite/r Gutachter: Prof. Dr. Levente Kriston

ABSTRACT

Background: Depressive disorders are among the most prevalent and most disabling disorders worldwide. Despite well-established diagnostic criteria and the availability of effective treatments, depressive disorders often remain undetected and untreated. Internet-based depression screening followed by automated feedback of the screening results could reach affected individuals outside the healthcare system and enhance early detection and treatment uptake. However, a thorough evaluation of the benefits and potential harms remains lacking.

Objective: The overarching objective of this cumulative dissertation is to evaluate the benefits and harms of an automated feedback after internet-based depression screening intervention in adults with suspected but undiagnosed depressive disorder. The dissertation synthesises four publications of the DISCOVER project and integrates all quantitative and qualitative findings within a mixed-methods framework.

Findings: Publication I presents the study protocol for the three-armed, observer-masked, randomised controlled DISCOVER trial (RCT). In total, 1178 individuals (aged ≥ 18 years, undiagnosed and untreated for depression) who screened positive on an internet-based self-report depression scale (Patient Health Questionnaire-9 ≥ 10 points) were randomised 1:1:1 to non-tailored, tailored, or no feedback of the screening results. Follow-up assessments were set at one and six months. Across all study arms, participants had a mean age of 37.1 years, 70% identified as female, and 62% met criteria for major depressive disorder (assessed by Structured Clinical Interview for DSM-5). Publication II reports that automated feedback did not significantly reduce depression severity compared with no feedback. Secondary outcomes, including depression-related health-behaviour, service uptake, and quality of life, also showed no significant improvements. Publication III indicates that feedback did not increase negative effects including misdiagnosis, mistreatment, or deterioration in symptoms and concern about symptoms. However, non-tailored feedback was potentially linked to worsening suicidal ideation. Publication IV presents a reflexive thematic analysis of interviews with 26 participants (conducted upon completion of the RCT) and highlights a two-step process underlying participants' experiences with the screening procedure across all study arms, i.e. irrespective of feedback provision. Step 1 describes the *recognition of depressive symptoms* ranging *from denial to awareness* as an initial reaction to particularly the screening questions. Step 2 describes a subsequent self-explorative process that encompasses up to three mutually

reinforcing themes: *cognitive positioning (rejection vs acceptance), emotional reactions (between empowerment and overload),* and *personal activation (from reflection to action).*

Integration of findings: The mixed-methods integration reveals that qualitative insights primarily complemented pre-defined outcomes and expanded the evaluation of automated feedback by also addressing depression screening alone. Regarding benefits, both depression screening questions and feedback of the results seemed to foster *self-awareness, self-acceptance*, and a *feeling of relatedness*. Regarding harms, findings suggest that feedback potentially triggered *symptom deterioration (increase in suicidal ideation)*, while screening questions contributed to *emotional burden* in some cases.

Conclusion: This dissertation concludes that automated feedback after internet-based depression screening does not yield relevant benefits regarding depression-related outcomes including behavioural patient activation, service uptake, or symptom reduction. On the other hand, feedback can potentially lead to harms such as increase in suicidal ideation. As such, feedback alone does not suffice to improve early detection and treatment uptake of affected individuals. However, both screening questions and feedback demonstrate potential to empower individuals by enhancing self-awareness, self-acceptance and relatedness. As answering depression screening questions may be emotionally burdensome for affected but undiagnosed individuals, there is a need for approaches that leverage the aforementioned empowering aspects of internet-based depression screening interventions while mitigating potential harms. Future research should corroborate the findings on patient-oriented outcomes and harms, ideally by using multicomponent intervention frameworks to disentangle differential effects of feedback vs screening alone. Further, practical efforts should prioritise patient safety through improved regulation and monitoring of existing internet-based screening services, e.g. by providing patient information and human guidance. Altogether, this dissertation highlights the absence of a straightforward solution for addressing undetected depression while underscoring the need for future research to better understand the pathway from identification to effective management of depressive disorders.

ZUSAMMENFASSUNG

Hintergrund: Depressive Störungen gehören zu den weltweit häufigsten und am stärksten beeinträchtigenden Erkrankungen. Trotz etablierter diagnostischer Kriterien und wirksamer Behandlungsansätze bleiben depressive Störungen häufig unentdeckt und unbehandelt. Internetbasiertes Depressions-Screening in Kombination mit automatisiertem Feedback der Screening-Ergebnisse könnte Betroffene erreichen und so deren Früherkennung und -behandlung verbessern. Eine umfassende Evaluierung der potenziellen Nutzen und Risiken dieses Ansatzes steht jedoch aus.

Ziel: Das übergeordnete Ziel dieser kumulativen Dissertation ist es, Nutzen und Risiken einer automatisierten Rückmeldung nach einem internetbasierten Depressions-Screening bei Erwachsenen mit vermuteter, undiagnostizierter depressiver Störung zu evaluieren. Die Dissertation basiert auf vier Publikationen des DISCOVER-Projekts und integriert deren quantitative und qualitative Ergebnisse innerhalb eines Mixed-Methods-Modells.

Ergebnisse: Publikation I präsentiert das Studienprotokoll für die dreiarmige, Beobachter-verblindete, randomisiert-kontrollierte DISCOVER-Studie (DISCOVER-RCT). Insgesamt wurden 1178 Personen (≥18 Jahre alt, undiagnostiziert und unbehandelt bezüglich einer depressiven Störung) 1:1:1 auf folgende Studienarme randomisiert: nicht-personalisiertes Feedback, personalisiertes Feedback oder kein Feedback zu den Screening-Ergebnissen. Die Follow-up-Erhebungen fanden nach einem und sechs Monaten statt. Das Durchschnittsalter der Teilnehmenden lag über alle Studienarme hinweg bei 37,1 Jahren, 70 % identifizierten sich als weiblich und 62 % erfüllten die Kriterien für eine Major Depression (erhoben mittels Strukturiertem Klinischen Interview nach DSM-5). Publikation II berichtet, dass das automatisierte Feedback die Depressionsschwere im Vergleich zu keinem Feedback nicht signifikant verringerte. Sekundäre Ergebnisse, einschließlich depressionsbezogenem Hilfesuchverhalten, Inanspruchnahme von Gesundheitsdiensten und Lebensqualität, zeigten ebenfalls keine signifikanten Verbesserungen. Publikation III deutet darauf hin, dass Feedback nicht zu negativen Effekten wie Fehldiagnosen, Fehlbehandlungen und der Verschlechterung von Symptomen oder Sorgen über die Symptome führte. Es wurde jedoch eine potenzielle Verbindung zwischen nicht-personalisiertem Feedback und einer erhöhten Suizidalität festgestellt. Publikation IV berichtet von einer reflexiven thematischen Analyse von Interviews mit 26 Teilnehmenden (durchgeführt nach Abschluss der RCT), deren Erfahrungen mit dem Screening-Verfahren über alle drei Studienarme hinweg im Rahmen eines zweistufigen Prozessmodells beschrieben wurden. Schritt 1 beschreibt das

(*An)Erkennen depressiver Symptome* (*von der Ablehnung bis zur Bewusstwerdung*) als erste Reaktion insbesondere auf die Screening-Fragen. Schritt 2 beschreibt einen darauffolgenden selbst-explorativen Prozess, der bis zu drei Themen umfasst: *kognitive Positionierung (Ablehnung vs. Akzeptanz), emotionale Reaktionen (zwischen Empowerment und Überforderung) und persönliche Aktivierung (von der Reflexion zur Handlung).*

Integration der Ergebnisse: Die Mixed-Methods-Integration zeigt, dass qualitative Erkenntnisse hauptsächlich die vordefinierten Endpunkte ergänzten sowie über das Feedback der Screening-Ergebnisse hinaus auch das Depressions-Screening selbst adressierten. So zeigt sich im Hinblick auf den Nutzen der Intervention, dass sowohl die Depressions-Screening-Fragen als auch das Feedback die *Selbstwahrnehmung*, *Selbstakzeptanz* und das *Gefühl der Zugehörigkeit* förderten. In Bezug auf potenzielle Risiken deuten die Ergebnisse darauf hin, dass Feedback potentiell zur Verschlechterung der Symptome führte und die Screening-Fragen zu erhöhter emotionaler Belastung beitrugen.

Schlussfolgerung: Diese Dissertation kommt zu dem Schluss, dass automatisiertes Feedback nach einem internetbasierten Depressions-Screening keinen relevanten Nutzen bzgl. depressionsbezogener Endpunkte wie Patientenaktivierung, Inanspruchnahme von Gesundheitsdiensten oder Symptomreduktion hat. Auf der anderen Seite birgt es potenziell das Risiko einer Symptomverschlechterung. Während Feedback allein nicht ausreicht, um die Früherkennung und -behandlung von Betroffenen zu verbessern, zeigen sowohl die Screening-Fragen als auch das Feedback Potenzial Betroffene subjektiv zu bestärken, indem sie ihre Selbstwahrnehmung, Selbstakzeptanz und ihr Gefühl der Zugehörigkeit fördern. Da die Screening-Fragen auch emotionale Belastung auslösen können, erscheinen Interventionen notwendig, die die bestärkenden Aspekte des Screening-Verfahrens adressieren und gleichzeitig negative Effekte minimieren. Zukünftige Forschung sollte die Befunde bzgl. Patienten-orientierter Endpunkte sowie Risiken untermauern, z.B. mittels Mehrkomponenten-Interventionen. Zudem sollte durch eine verbesserte Regulierung bestehender internetbasierter Depressions-Screening-Angebote die Patientensicherheit gefördert werden. Ansatzpunkte könnten die Begleitung durch Ansprechpersonen und verbesserte Patienteninformationen zum Nutzen und Risiko entsprechender Dienste sein. Insgesamt betont diese Dissertation die Notwendigkeit zukünftiger Forschung, um den komplexen Weg von der Identifikation bis zum effektiven Management depressiver Störungen besser zu verstehen.

TABLE OF CONTENT

LIST OF ABBREVIATIONS

APA	American Psychological Association
CI	confidence interval
DSM-5	Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
e.g.	exempli gratia (for example)
et al.	et alia (and others)
EQ-5D-5 L VAS	EuroQoL-5 Dimensions-5 Level visual analogue scale
GAD-7	Generalized Anxiety Scale-7
ICD-11	International Classification of Diseases, Eleventh Edition
i.e.	id est (that is)
ITT	intention-to-treat
PHQ-9	Patient Health Questionnaire-9
RCT	randomised controlled trial
RR	relative risk
SCID	Structured Interview for DSM
SSS-8	Somatic Symptom Scale-8
WHO	World Health Organization

1 INTRODUCTION AND BACKGROUND

> "Evidence has accumulated over decades that depression is a leading cause of avoidable suffering in the world. Yet, too few people […] understand or acknowledge depression as distinct from the other troubles that people face."
>
> Lancet–World Psychiatric Association Commission (Herrman et al., 2022)

Worldwide and in Germany, depressive disorders range among the most prevalent and most disabling disorders (Jacobi et al., 2014; World Health Organization [WHO], 2022). However, and despite clear diagnostic criteria and the availability of effective treatments, depressive disorders often remain undetected and therefore untreated (Vigo et al., 2020). An intervention consisting of internet-based depression screening combined with automated feedback of the screening results may promote early detection and treatment of those affected but yet undetected - i.e. those who are potentially not aware that their condition is 'distinct from the other troubles they face'. This dissertation aims at comprehensively evaluating potential benefits and harms of an automated feedback after internet-based depression screening intervention developed and tested at the Department of Psychosomatic Medicine and Psychotherapy of the University Medical Center Hamburg-Eppendorf. The dissertation is cumulative in nature and is based on four publications. In the following, I will explicate the relevant background regarding depressive disorders and their under-detection, the rationale and current scientific evidence of feedback after internet-based depression screening as an approach to improve early detection, and the overall research question addressed in this work. In Section 2, I will outline the underlying research project DISCOVER and specify the mixed methods approach of the present evaluation. The four publications reporting on procedures and findings of the DISCOVER trial will be summarised in Section 4, followed by an integration of quantitative and qualitative findings in Section 5. Lastly, in Section 6, I will discuss overall findings against the background of potential study limitations and draw conclusions on implications for research and practice in the field of early detection of depressive disorders. The complete publications are listed in Section 7.

1.1. Depressive disorders

1.1.1. Definition and classification

Suffering that resembles the modern definition of depression has been described throughout human history for thousands of years. Often associated terms such as "melancholy" or "poor spirits" indicate the difficulty of distinctly defining a condition that overlaps with normal human response to adversity (Goldberg, 2011; Murphy, Laird, Monson, Sobol, & Leighton, 2000). Indeed, depressive disorders encompass a heterogeneous syndrome, with different combinations of unspecific symptoms that produce diverse clinical phenotypes (Fried & Nesse, 2015; Goldberg, 2011). In science and medical practice, classification systems such as the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5; American Psychiatric Association, 2013) and the International Statistical Classification of Diseases and Related Health Problems, 11th revision (ICD-11; World Health Organization, 2019) conceptualise depressive disorders as syndromes characterised by states of low mood or loss of pleasure or interest, accompanied by further mental and physical symptoms such as poor sleep, poor appetite or decreased activity, that are serious enough to impair functioning in social, occupational or other situations. The spectrum of depressive disorders comprises major depressive disorder, characterised by single or recurrent episodes persisting for at least two weeks, and chronic forms that persist for at least two years and are characterised by milder symptoms (American Psychiatric Association, 2013).

1.1.2. Aetiology and course

Current aetiological approaches are multifactorial and view depressive disorders as the result of an inter-individually varying combination of factors. Often, negative live events act as triggers for the onset of an episode and typically interact with a unique interplay of genetic, environmental, social, and developmental vulnerability and resilience factors (Berger, van Calker, Brakemeier, & Schramm 2019; Herrman et al., 2022). The temporal course of depressive disorders is similarly diverse. Many depressive episodes remit within one year, between 12% and 61% of affected individuals are estimated to experience episodes lasting more than one year, and an estimated proportion of 27% to 85% suffers from an intermittent recurrence over the life course (see Herrman et al., 2022, for an unsystematic review). The risk of recurrence increases with every episode. It is estimated to range between 40% and

60% after the first episode and can rise to approximately 90% after the third and subsequent episodes (Monroe & Harkness, 2011).

1.1.3. Prevalence and burden

As of 2019, an estimated 280 million (3.8%) of people worldwide lived with a depressive disorder, making it the second leading cause of disability worldwide (WHO, 2022). Estimates on prevalence vary by age (5% among adults and 5.7% among adults older than 60 years), sex (6% among females and 4% among males), and world region (GBD 2019 Mental Disorders Collaborators, 2022). For Germany, for example, a representative population-based study reported a twelve-month prevalence of 7.7%, which corresponds approximately 4,9 million German individuals (Jacobi et al., 2014). The prevalence of a lifetime physician's diagnosis of a depressive disorder is reported to vary between 11.6% and 15,9% (Jacobi et al., 2014; Streit et al., 2023). Depressive disorders substantially affect the individuals' overall health and well-being. They are ranked among the leading causes of non-fatal health loss globally (GBD 2019 Mental Disorders Collaborators, 2022), often co-occur with and potentially complicate somatic and other mental disorders (Gold et al., 2020; Steffen, Nübel, Jacobi, Bätzing, & Holstiege, 2020), and are associated with an increased mortality through increased risk for somatic disorders and suicide (Chesney, Goodwin, & Fazel, 2014). Depressive disorders also have a high societal impact on the population level; they are associated with increased health service utilisation, decreased work productivity, increased burden on family members (Lim, Jacobs, Ohinmaa, Schopflocher, & Dewa, 2008), and eventually high economic costs (König, König, & Konnopka, 2020). In Germany, for example, depressive disorders are responsible for the highest number of days of sick leave, with numbers being on the rise (Baumeister et al., 2015; Schneider, Erhart, Hewer, Loeffler, & Jacobi, 2019)

1.1.4. Treatment

First-line treatments recommended for routine clinical care of depressive disorders are antidepressants, psychotherapy, and a combination of both (Bundesärztekammer [BÄK], Kassenärztliche Bundesvereinigung [KBV], & Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften [AWMF], 2022); WHO, 2015). There are also digital health applications, which are based on effective psychotherapeutic approaches but are designed to be entirely or mostly self-delivered. A solid meta-analytical evidence base indicates that both antidepressants (Cipriani et al., 2018) as well as psychotherapy and digital

health applications are moderately effective with regard to symptom reduction, response, and remission (Cuijpers et al., 2023; Karyotaki et al., 2021). In Germany, effective treatments such as evidence-based antidepressants, psychotherapy, and certified digital health applications are covered by statutory health insurances (see Schreiter et al., 2023, for a scoping review on digital health applications prescribable in Germany).

1.2. Early depression detection via depression screening

If left untreated, individuals with depressive disorders have an increased likelihood of chronification, treatment resistance, and an increased disease burden (Kraus, Kadriu, Lanzenberger, Zarate, & Kasper, 2019). Timely referral to effective treatments, in turn, requires the identification and diagnosis of depressive disorders early in their course. Therefore, early detection is considered an essential step in depression care (Cacheda, Fernandez, Novoa, & Carneiro, 2019).

1.2.1. Under-detection of depressive disorders

However, despite clear diagnostic criteria and the availability of effective treatments, depressive disorders often remain undetected and therefore untreated. For example, data from the 2015 WHO World Mental Health Survey covering 15 countries indicate that among adults meeting criteria for a major depression, on average less than 50% had any contact with a health care specialist, and only about 10% received evidence-based care (Vigo et al., 2020). Numbers in Germany are higher, but similarly mirror a gap in early detection and treatment. In a cross-sectional epidemiological study conducted in primary care in 2017, of patients meeting criteria for a major depression only 50% received a respective diagnosis from their physician, and only 40% were treated in conformity with current depression guidelines (Beesdo-Baum et al., 2018; Trautmann & Beesdo-Baum, 2017). Discussed reasons for the under-detection of depressive disorders include structural problems such as time restrictions and lacking reimbursements in primary care. Additionally, identification is complicated by the heterogeneity of clinical phenotypes and under-reporting of mental symptoms, which is partly associated with the stigma of depressive disorders (Beesdo-Baum et al., 2018; Heinz et al., 2021; Kratz et al., 2003).

1.2.1. Definition and current practice of depression screening

Depression screening can be a proactive and straightforward approach to address barriers related to the detection and under-reporting of depressive disorders. Depression

screening is mostly defined as administering a depression symptom questionnaire to individuals in health care settings who are not already diagnosed, and then classifying positive and negative results based on a pre-specified cut-off score. In the following, screened individuals with positive results should be provided with comprehensive mental health assessments, and, as indicated, further depression management (Thombs, Markham, Rice, & Ziegelstein, 2021). Depression screening in routine clinical care, either for all individuals or for at-risk populations, is often recommended in nationally acknowledged guidelines on depression care such as the 2023 US Preventive Services Task Force guideline (Barry et al., 2023) and the 2022 German National Disease Management guideline (BÄK, KBV, & AWMF, 2022).

1.2.2. Empirical evidence on depression screening

With regard to diagnostic accuracy, two meta-analyses agree that well-established screening tools for depression, e.g. the Patient Health Questionnaire-9 (PHQ-9, Kroenke, Spitzer, & Williams, 2001), can accurately identify adults with depressive disorders in primary care or comparable settings (see Negeri et al., 2021, for an individual-participant-meta-analysis on the PHQ-9; and O'Connor et al., 2023, for a meta-analysis on the four most widely used tools). Current evidence with regard to the efficacy regarding health outcomes, however, is limited. Based on two recent systematic evidence reports it can be concluded that screening alone is insufficient, as positive effects on depression-related outcomes depend on subsequent care management components, i.e. a post-screening referral to depression-related diagnostics and evidence-based treatment (Beck et al., 2022; O'Connor et al., 2023).

1.3. Automated feedback after internet-based depression screening

1.3.1. Rationale for providing feedback of depression screening results

One promising and low-threshold approach to stimulate a post-screening referral process is to provide feedback of the depression screening results directly to the individual. The rationale of this approach follows self-regulation theories of health behaviour such as the Common Sense Model of health and illness behaviour (Leventhal, Meyer, & Nerenz, 1980). According to this theory, functional illness behaviour aimed at managing a condition depends on the individual's subjective model of the illness, i.e. in how the individual perceives and interprets the symptoms, as well as on the individual emotional response to the symptom experience. Specifically, the theory postulates that individuals undergo a self-diagnostic process in which they form beliefs about the symptoms' belonging to a diagnosis, their cause,

timeline, consequences, and controllability. In this framework, delay in self-referral is mainly attributed to misappraisal of symptoms as not belonging to an illness and/or as not requiring care (Martin, Rothrock, Leventhal, & Leventhal, 2003). Indeed, in terms of depressive disorders, affected individuals often fail to connect their symptoms to the diagnosis, i.e. to define their symptoms as a depressive disorder (see Doblyte & Jimenez-Mejias, 2017, for a qualitative synthesis). Providing affected individuals with feedback on their screening results could help these individuals recognise that they have depressive symptoms, link them to adequate resources, and empower them in evidence-based decision making. This, in turn, could motivate a functional management of their condition, including the uptake of adequate depression care. Empirical evidence from two randomised controlled trials conducted by our research group corroborates this rationale: In cardiology, a printed, patient-targeted feedback of the depression screening results together with the advice to seek professional diagnostic advice stimulated an increased search for information about depression and lead to a reduction in depression severity after six months (Löwe et al., 2016). Similarly, in primary care, an elaborated version of this patient-targeted feedback intervention improved the subjective patient-physician communication about depression and increased the likelihood of being offered psychotherapy, compared to feedback to the general practitioner only or no feedback. Subgroup analyses further showed that women, those with a history of lifetime depression, and those without addiction might experience beneficial effects of feedback also with regard to depression severity (Löwe et al., 2024).

1.3.2. Rationale for an internet-based implementation

In the past decades, leveraging digital technologies has become an important new frontier in mental health care. The internet has developed to one of the primary sources for information on mental health (Eichenberg, Wolters, & Brähler, 2013; Vaidyanathan et al., 2022), and digital approaches aim to extend the reach of clinical interventions to underserved populations (Andersson, 2018). Internet-based depression screening is already promoted by mental health care providers and frequently used by those seeking diagnostic advice. In the United States, for example, each year approximately 1 million people use the online depression screening tool of the leading non-profit mental health organisation Mental Health America (Yom-Tov et al., 2023). Similarly, in Germany, each year half a million users complete the online depression screening tool on the website of Stiftung Deutsche Depressionshilfe, one of the leading German patient self-help organisations (Stiftung Deutsche Depressionshilfe und Suizidprävention, 2024, unpublished data from personal

communication). A study by Yom-Tov et. al further suggests that online screening is particularly prominent in underserved rural communities (Yom-Tov et al., 2023). Against the background of wide public interest in internet-based depression screening, translating our evidence-based feedback intervention to the internet could be a promising approach to reach those affected but not yet reached by the health system. Additionally, it allows to directly link internet-based information and self-help resources such as digital health applications. Lastly, automated feedback could improve the salience and fit of information by tailoring the feedback to the characteristics of the individual.

1.3.3. Empirical evidence on benefits and harms

Widely accessible screening for depression, as for any other health condition, should only be implemented after a careful and systematic evaluation of potential benefits and harms (see Dobrow, Hagens, Chafe, Sullivan, & Rabeneck, 2018, for a systematic review and consensus paper on screening principles). Despite the outlined potential and widespread use of internet-based depression screening followed by automated feedback, current empirical evidence on benefits and harms is limited and inconclusive. With regard to benefits, two observational studies report service uptake rates after internet-based feedback ranging between 30% and 60% (BinDhim et al., 2016; Choi et al., 2018). More rigorous research, on the other hand, has failed to confirm positive effects on help-seeking. In the only published randomised controlled trial on feedback after internet-based depression screening, feedback (vs a generic advice to seek help) of screening results had no significant effect on professional help-seeking three months later (Batterham, Calear, Sunderland, Carragher, & Brewer, 2016). However, this trial was not sufficiently powered, had a high attrition rate associated with receiving the feedback intervention, did not exclude cases who were already diagnosed and/or treated, and had a short follow-up period of three months. Regarding subjectively perceived benefits of screening and feedback, one focus group study with young adults suggests that internet-based depression screeners meet individual emotional needs for validation and self-understanding (Kruzan et al., 2022). Although restricted to a very young population and not excluding cases already diagnosed, this study expands on quantitative findings by highlighting the potential subjective value of the screening process itself. Regarding harms of automated feedback after internet-based depression screening, the scientific debate is limited to opinion papers or unsystematic evidence. Discussed negative effects include inadequate management and care for individuals who screened false-positive, negative psychological effects such as distress, stigma, or nocebo effects such as deterioration of symptoms

(Danczak, 2017; Duckworth & Gilbody, 2017; Ryan & Wilson, 2008; B. Thombs, Turner, & Shrier, 2019). Indeed, in the aforementioned qualitative study on internet-based mental health screening in young adults some participants described having been discouraged, shocked or concerned by the feedback (Kruzan et al., 2022). Further, an observational study found that internet-based screening procedures that included referrals to in-person care in their feedback had a higher likelihood of subsequent online searches with suicidal intent, potentially suggesting an increase of suicidal ideation (Jacobson et al., 2022). However, robust systematic research on these potential negative effects is missing.

1.4. Aim of this dissertation

Depressive disorders are a prime target for early detection, as is shown by their high prevalence, disease burden, and recurrence if left untreated (Section 1.1.) and the high rates of under-detection and under-treatment despite available and effective treatments (Section 1.2.). Internet-based depression screening followed by automated feedback of screening results may be promising to reach affected but undiagnosed individuals. To date, however, a thorough evaluation of benefits against the background of potential harms is outstanding (Section 1.3.).

In response to this gap, the present dissertation evaluates an automated feedback after internet-based depression screening intervention aiming at improving early detection and treatment uptake of individuals with suspected but undiagnosed depressive disorder. Drawing on four publications, this dissertation integrates findings on the efficacy and potential negative effects of automated feedback after internet-based depression screening with qualitative insights into participants' experiences with the whole screening process. The overarching research question guiding this work is:

What are the benefits and harms of an automated feedback after internet-based depression screening intervention in adults with suspected but undiagnosed depressive disorders?

2 METHOD

2.1. The DISCOVER project

This dissertation is embedded in the research project DISCOVER that aimed at testing the efficacy of two versions of automated feedback after internet-based depression screening in adults with suspected but undiagnosed major depressive disorder in a three-armed, randomised controlled trial (RCT). During the course of the study, the project was extended by a qualitative interview study on participants' experiences with the screening process. The project was funded by the German Research Foundation and conducted at the Department of Psychosomatic Medicine and Psychotherapy of the University Medical Center Hamburg-Eppendorf between April 2020 and September 2023 (principal investigator Prof Dr Sebastian Kohlmann). The research team comprised a clinical psychologist, a medical doctor, two health economists, and three biometricians, all with extensive experience in intervention research, as well as a doctoral candidate in clinical psychology. Procedures involved in the underlying studies have been approved by the independent ethics committee of the Hamburg Medical Chamber (Ärztekammer Hamburg; July 2019, reference number PV7039) and the ethics committee of the University Medical Center Hamburg-Eppendorf (June 2021, reference: 0337). The DISCOVER trial was pre-registered at ClinicalTrials.gov (Nov 2020, identifier NCT04633096) and quantitative main and secondary analyses were pre-specified before analysing the respective data (https://osf.io/mnqvs/). Conducting and reporting of all resulting publications were in accordance to appropriate CONSORT 2010 extensions (Daniela et al., 2023; Eysenbach, 2013; Montgomery et al., 2018) or the COREQ checklist (Tong, Sainsbury, & Craig, 2007). Study data and a statistical report with unpublished additional analyses are available upon request from the principal investigator.

2.2. Design of this dissertation

In order to address the overarching research question, this dissertation cumulates four publications generated within the DISCOVER project and focusses on integrating the qualitative and quantitative findings within an overall mixed methods evaluation. Given the complexity of depression management, a mixed-methods framework appears suitable to provide a nuanced evaluation of both benefits and harms. The evaluation is conceptualised within a paradigmatic framework of pragmatism, which enables the integration of different epistemological perspectives. As such, the evaluation can combine both post-positivist,

quantitative methods and interpretivist, qualitative insights, creating a more comprehensive understanding of the subject (Dawadi, Shrestha, & Giri, 2021). Within this pragmatic framework, the DISCOVER project can be described as following an embedded mixed methods interventional design (Fetters, Curry, & Creswell, 2013), with a qualitative interview study being embedded within an RCT. For the RCT, eligible participants screening positive for major depression on the internet-based PHQ-9 (Kroenke et al., 2001) were randomly assigned 1:1:1 to receive automated non-tailored, tailored, or no feedback on their screening results. Outcome assessments were set at one and six months after screening. After having completed the quantitative data collection, a subsample of participants underwent semi-structured interviews. Qualitative data were analysed and published prior to the commencement of quantitative data analysis. The overall integration of findings for this dissertation was carried out at the stage of reporting and interpreting the results (see Figure 1). Conducting and reporting of this mixed-methods evaluation follows the journal article reporting standards for qualitative primary, qualitative meta-analytic, and mixed methods research in psychology (Levitt et al., 2018).

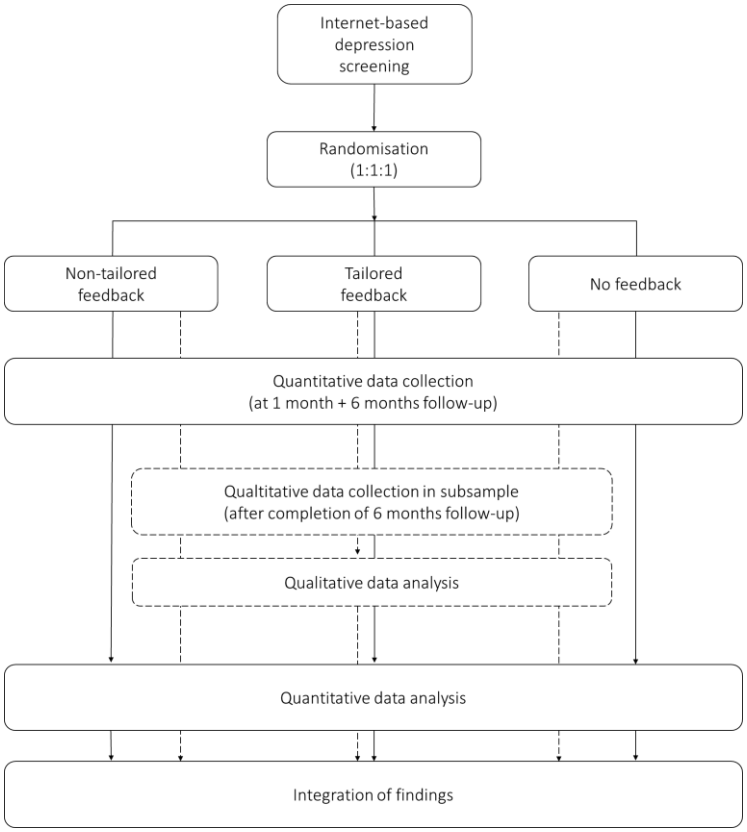


Figure 1. Embedded mixed methods interventional design of the DISCOVER project.

2.3. Underlying Publications

All publications, along with their supplementary materials, are listed in Section 7. Summaries of the publications are provided in Section 3. Publication I is a study protocol of the DISCOVER trial. It transparently describes trial procedures, the feedback interventions, and outcome measures. Publication II reports on the efficacy of automated feedback after internet-based depression screening with regard to depression severity (primary outcome), service uptake, and further clinical outcomes, as compared to no feedback. Publication III reports on potential negative effects associated with the feedback (secondary analyses), as compared to no feedback. Outcomes relate to misdiagnosis, mistreatment, symptom deterioration, and deterioration in emotional response to symptoms. Publication IV describes how a subsample of participants experienced the screening process (qualitative analysis of interview data). In this study, participants' reports are summarised in four themes alongside a two-step process and include subjectively perceived positive and negative effects of the internet-based screening procedure (not differentiating between feedback and no feedback).

2.4. Integration of findings

The findings of publications II to IV were integrated to address the overarching research question by determining their fit alongside overarching harms and benefits. The overarching categories of harms and benefits were built inductively from the findings: First, qualitative and quantitative findings regarding benefits and harms were mapped against each other. Second, overarching themes were developed by clustering matching findings around a central organising concept. Lastly, overarching themes were named. Eventually, the fit between qualitative and quantitative findings within one overarching theme was interpreted with regard to convergence, divergence, complementarity, or expansion, resulting in integrated interpretations called meta-inferences (Fetters, 2019). Convergence and divergence describe the agreement or disagreement between the findings, complementarity refers to findings that illustrate different but non-contradictory interpretations, and expansion occurs when some findings overlap but also provide space for further interpretation. The results of this integration process are reported using a narrative side-by-side joint display (Guetterman, Creswell, et al., 2015).

3 SUMMARIES OF PUBLICATIONS

In the following, I summarise the four publications underlying this dissertation. As the overarching theoretical background is already detailed in Section 1, the summaries start with the specific objective of the respective publication. Additionally, as the publications build on each other, redundant content regarding trial procedures is omitted to avoid repetition.

3.1. Publication I: Study protocol

Sikorski, F., König, H.-H., Wegscheider, K., Zapf, A., Löwe, B., & Kohlmann, S. (2021). The efficacy of automated feedback after internet-based depression screening: Study protocol of the German, three-armed, randomised controlled trial DISCOVER. *Internet Interventions*, 25, 100435.

3.1.1. Objective

This study protocol describes the design and procedures of the DISCOVER trial.

3.1.2. Method

DISCOVER is an internet-based, observer-masked, three-armed, randomised controlled superiority trial. It is designed to test the efficacy of two versions of automated feedback after internet-based depression screening with regard to the change in depression severity six months after screening. The trial is promoted as a financially compensated 'online study on stress and psychological well-being', with recruitment being conducted nationwide in Germany through traditional and social media and a nationwide online access survey panel. Powered to detect a small mean group difference and accounting for an estimated dropout of 35%, the study aims to recruit a total of 1074 participants who screen positive for depression on the PHQ-9 (PHQ-9 ≥ 10 points). Participants must be aged ≥18 years and be undiagnosed/untreated for depression within the last year. After completing eligibility assessment and depression screening, participants are randomly allocated 1:1:1 to receive no feedback, automated non-tailored feedback, or automated tailored feedback on the depression screening result. Randomisation is conducted electronically, based on permuted block technique, and stratified by depression severity (moderate vs severe). Outcome assessments are set at one month and six months after screening and are organised electronically, with participants being automatically reminded in case of non-respondence. To verify the presence

of a major depression, the Structured Clinical Interview for DSM-5 (SCID; Beesdo-Baum, Zaudig, & Wittchen, 2019) is conducted via telephone at baseline and six months after screening. Research staff is masked to group allocation and outcome assessment as this is done electronically; participants cannot be masked but are kept unaware of trial hypotheses to minimise expectancy bias.

The non-tailored feedback includes the depression screening result, a recommendation to seek professional diagnostic consultation, and brief general information on depression and its treatment including hyperlinks to health or social services (e.g. to directly make an appointment with a health care specialist or to digital health applications, see Figure 2 or supplementary Figures in section 7.2.2). The tailored feedback includes the same basic information, but is individually framed according to the participants' symptom profiles, treatment preferences, causal symptoms attributions, health insurance, and local residence (see supplementary Figures in section 7.2.2). The feedback interventions were developed in a multistage process together with patient representatives. To ensure participants' safety, all participants who indicate elevated suicidal ideation are additionally shown a screen providing an advice to urgently seek help together with information on available help services.

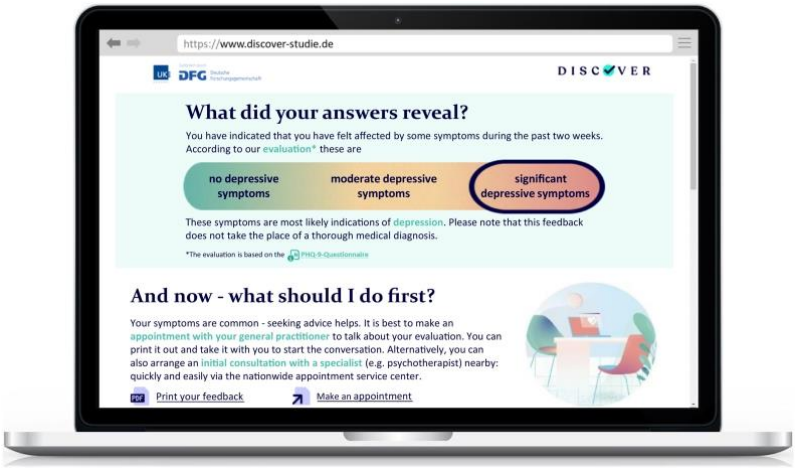


Figure 2. First screen of non-tailored automated feedback of internet-based depression screening results (English translation).

The primary outcome is the change in PHQ-9 score six months after screening, with total scores ranging from 0 to 27 and higher scores indicating greater depression severity.

Secondary outcomes include the number of participants receiving evidence-based depression care (i.e. psychotherapy and/or antidepressant treatment) since study start, the number of participants reporting to have received a diagnosis of a depressive disorder by a health care professional since study start, the number of participants reporting to have engaged in depression-related health behaviour since study start, health-related quality of life (EuroQoL-5 Dimensions-5 Level visual analogue scale [EQ-5D-5 L VAS]; Gunther, Roick, Angermeyer, & König, 2008), anxiety severity (Generalized Anxiety Disorder-7 [GAD-7]; Lowe et al., 2008), and somatic symptom severity (Somatic Symptom Scale-8, SSS-8 [SSS-8]; Gierk et al., 2014). To estimate possible unintended negative events, the occurrence of any negative event attributed to trial participation is assessed six months after randomisation with an open question via telephone.

3.1.3. Strengths and limitations

Strengths of the DISCOVER trial include the validation of the PHQ-9 screening results with a semi-structured interview, elaborated development of the feedback interventions, as well as advantages of the internet-based implementation such as automation of assessments to improve retention, mechanisms to ensure the singularity and validity of the data, and well-designed online questionnaire administration to minimise participant burden. Limitations include the self-selection of the sample and potential confounding of feedback effects with the screening effects, which is not controlled for.

3.1.4. Discussion

The DISCOVER RCT is well designed to yield comprehensive information on how automated feedback after internet-based screening could improve early detection and resolution of depression. If proven efficacious, the low-threshold feedback intervention could be easily and widely implemented in various settings.

3.2. Publication II: Efficacy evaluation

Kohlmann, S., Sikorski, F., König, H.-H., Schütt, M., Zapf, A., Löwe, B., (2024). The efficacy of automated feedback after internet-based depression screening (DISCOVER): an observer-masked, three-armed, randomised controlled trial in Germany. *The Lancet Digital Health*, *6*(7), e446-e457.

3.2.1. Objective

This publication reports the main findings of the DISCOVER RCT. The primary hypothesis was that automated feedback (irrespective of the feedback mode) leads to a greater reduction in depression severity six months after screening compared with no feedback. The secondary hypothesis was that tailored feedback leads to a greater reduction in depression severity than non-tailored feedback. Additionally, we exploratively investigated the effects of feedback on the initiation of depression care and depression-related health behaviour, as well as other clinical outcomes six months after screening.

3.2.2. Method

The design, procedures, and outcome measures of the trial are described in Section 3.1.2. Analyses were conducted as described in the pre-registered statistical analysis plan (see supplementary material in Section 7.2.2).

Hypotheses were two-sided and tested for differences at a 5% significance level. The primary analysis was performed as an ANCOVA of the PHQ-9 change scores (baseline to 6-months follow-up), with the baseline value as a covariate. According to the closed testing principle, subsequent pairwise comparisons where conducted only in case of a significant overall *F* test of study arm. The analysis was done in the full analysis set population following the intention-to-treat (ITT) principle, i.e. including all participants as randomly assigned, provided they had valid baseline and 6-month PHQ-9 scores. To account for missing values, sensitivity analyses based on the last observation carried forward and multiple imputation approach were performed. An additional sensitivity analysis was conducted on the per-protocol population, excluding participants with protocol violations (e.g. not receiving the feedback, participating multiple times, or completing the baseline survey in under 2 minutes). Lastly, subgroup analyses were conducted to examine differences based on baseline

depression severity (moderate vs severe) and to adjust for fulfilling the DSM-5 based criteria for major depressive disorder and the subjective belief of having a depressive disorder.

Exploratory secondary outcome analyses were done in the respective full analysis set populations and for descriptive purposes only ($\chi 2$ test of independence for binary outcomes, and ANCOVA as defined for the primary outcome for continuous endpoints).

3.2.3. Results

Between Jan 12, 2021, and Sept 30, 2022, of 5457 participants, 4878 (89%) completed internet-based screening. Of these, 1178 (24%) screened positive for depression (PHQ-9 $\geq$ 10 points) while being undiagnosed and untreated. These were assigned to receive no feedback (n = 391), non-tailored feedback (n = 393), or tailored feedback (n = 394) on the screening result. Upon completion of data collection on Sept 30, 2022, 965 (92%) participants provided 6 months follow-up data on the PHQ-9 (see Figure 2 in Section 7.2.1 for a flow chart).

Participants were mostly recruited via social media (27%), an online access panel (14%), and search engines (10%). Sample characteristics were well balanced across study arms; the total mean age was 37.1 (standard deviation [SD] 14.2) years, 70% of participants were female, and 49% had a high education level. The average PHQ-9 depression severity score was 14.8 (SD 4.0), and 86% of participants thought that they might currently suffer from a depressive disorder. Of 909 participants interviewed with the SCID, 62% fulfilled the criteria for major depressive disorder. Regarding the use of the non-tailored and tailored feedback interventions, 95% and 94% participants opened the feedback screen, 34% and 36% interacted with the feedback intervention by clicking on information boxes or hyperlinks, and 59% and 58% downloaded the feedback, respectively.

Six months after screening, depression severity decreased by 3.4 to 3.7 PHQ-9 points (0.67 $\leq$ Cohen's $d \leq$ 0.74) across the three study arms, with no significant difference between study arms (p = 0.72, see Table 1). The results remained consistent in predefined sensitivity analyses (ps $\geq$ 0.47) and when statistically adjusting for baseline depression severity, fulfilment of the DSM-5 based criteria for major depressive disorder, and the subjective belief of having a depressive disorder. Regarding secondary outcomes, there were no relevant group differences in the number of participants initiating evidence-based depression care or depression-related health-behaviour, nor regarding other clinical outcomes (see Table 1).

Regarding negative events, one participant in the no feedback arm, four participants in the non-tailored feedback arm and four participants in the tailored feedback arm reported six

months after screening that trial participation was emotionally burdensome, associated with distressing memories, or associated with a feeling of helplessness.

Table 1. Primary and secondary efficacy outcomes six months after randomisation (full analysis set).

	Adjusted mean difference (95% CI)* or n (%)						p value
	n	No Feedback	n	Nontailored Feedback	n	Tailored Feedback	
Primary outcome							
Change in depression severity (PHQ-9)	325	-3·4 (-4·0 to -2·9)	319	-3·5 (-4·0 to -3·0)	321	-3·7(-4·3 to -3·2)	0·7190
Secondary outcomes							
Evidence-based depression care							
Diagnosis by a health care professional	324	43 (13%)	317	52 (16%)	320	53 (17%)	0·4267
Psychotherapy and/or antidepressant	325	82 (25%)	319	93 (29%)	321	91 (28%)	0·4994
Depression-related health behaviour							
Seeking information+	322	176 (54%)	317	171 (54%)	319	180 (56%)	0·8104
Seeking social support§	325	213 (66%)	319	201 (63%)	321	219 (68%)	0·3812
Self-management#	325	193 (59%)	319	197 (62%)	321	215 (67%)	0·1249
Seeking formal help$	325	134 (41%)	319	143 (45%)	321	156 (49%)	0·1700
Change in quality of life (EQ-5D-5L VAS)	321	4·1 (1·9 to 6·4)	312	3·9 (1·6 to 6·2)	318	3·2 (1·0 to 5·5)	0·8502
Change in anxiety severity (GAD-7)	323	-3·3 (-3·8 to -2·9)	314	-3·1 (-3·6 to -2·6)	318	-3·4 (-3·9 to -3·0)	0·6486
Change in somatic symptom severity (SSS-8)	322	-2·6 (-3·2 to -2·1)	314	-2·2 (-2·8 to -1·7)	318	-2·5 (-3·1 to -2·0)	0·5505

P values refer to F tests (continuous outcomes) or Chi-Square-tests (binary outcomes) and are not adjusted for multiple testing. CI=Confidence Interval. PHQ-9=Patient Health Questionnaire-9. + seeking information included information obtained through personal contact, print media or the internet. § seeking social support included contact with peers, friends, family or self-help groups. # self-management included increasing physical activity, using relaxation techniques, improving sleep hygiene, or using unguided self-help programmes (books or internet-based). $ seeking formal help included seeking contact with primary care physicians or mental health professionals. EQ-5D-5L=EuroQoL-5 Dimensions-5 Level scale. VAS=Visual analogue scale. GAD-7=Generalized Anxiety Disorder-7. SSS-8=Somatic Symptom Scale-8. *Change from baseline to 6-months follow-up; adjusted for corresponding outcome at baseline.

Table 2: **Primary and secondary outcomes six months after randomisation (full analysis set population)**

3.2.4. Strengths and limitations

Strengths of the trial are the large sample size, the lower-than-estimated loss-to-follow-up rate of 18%, and the diverse recruitment across Germany which should have ensured a sample that is representative of Germans interested in mental health. Further, the design enabled to disentangle the effect of screening and automated feedback from further depression care and included only the target group of individuals affected but not diagnosed or treated. Limitations include that the internet-based study inclusion and ”validity” of participants could not be verified in person, outcome data were self-reported, and the study did not explicitly call for participants seeking information on depression, who might have been more eager to follow the advice of the feedback.

3.2.5. Conclusion

This is the first trial that provides empirically robust evidence that neither non-tailored nor tailored automated feedback after internet-based depression screening improve relevant depression outcomes such as depression severity or the initiation of depression-related health

behaviour or care. Possible explanations for this negative finding refer to participants' lower than expected interaction with the feedback, the low-threshold nature of the one-time feedback, and the overall high level of engagement in depression-related health behaviour irrespective of the intervention. The findings should be critically considered by health care providers offering internet-based depression screening and by guideline developers recommending population-based depression screening.

3.3. Publication III: Negative effects evaluation

Sikorski, F., Löwe, B., Daubmann, A., & Kohlmann, S. (under review). Does feedback after online depression screening cause harm? A secondary analysis of negative effects in the randomised controlled DISCOVER trial. *Journal for Medical Internet Research.*

3.3.1. Objective

This manuscript reports secondary findings of the DISCOVER RCT, which refer to negative effects of automated feedback after internet-based depression screening. Specifically, it aims at examining whether the feedback is associated with misdiagnosis and mistreatment six months after screening, as well as deterioration in depressive symptoms, deterioration in emotional response to symptoms, and deterioration in suicidal ideation one and six months after screening.

3.3.2. Method

This secondary analysis uses data from the randomised controlled DISCOVER trial, described in Sections 3.1. and 3.2. Analyses were pre-registered after study initiation but before initiation of main analyses (https://osf.io/tzyrd).

Mistreatment and misdiagnosis at six months were assessed for participants who were screened false positive at baseline while not meeting DSM-5 criteria for major depression in the SCID interview performed post-screening. Misdiagnosis was operationalised as having received a depression diagnosis by a health professional since screening. Mistreatment was operationalised as having started psychotherapy or antidepressant medication since screening. Based on the statistically reliable change index, deterioration in depressive symptoms was defined as a pre-post change of ≥ 4.4 points in the PHQ-9, and deterioration in emotional response to symptoms as a pre-post change of ≥ 3.1 points in a composite scale based on the Brief Illness Perception Questionnaire (total scores ranging from 0 to 10). Deterioration in suicidal ideation was defined as a pre-post change of ≥ 1 point in the PHQ-9 suicide item (total scores ranging from 0 to 3). Analyses were conducted in the per-protocol sample, including participants with respective valid baseline and postbaseline values and no major protocol violation. Protocol deviations were predefined as not receiving the intervention, participating multiple times, indication at baseline that the survey was not seriously answered, completing the baseline survey in under 2 minutes, or providing an invalid email address.

Analyses were conducted using modified log-Poisson regressions with a robust sandwich variance estimator, resulting in relative risks (RR) for non-tailored and tailored feedback compared with no feedback, respectively. As sensitivity analyses, analyses were repeated in the ITT samples without and with imputed missing data (based on best case and worst case scenarios), and as logistic regressions, resulting in odds ratios.

3.3.3. Results

Out of 1178 participants randomised, 948 (81%) were included in the per-protocol sample, resulting in 312 (1 month) and 309 (6 months) analysed participants in the no feedback arm, 300 (1 month) and 296 (6 months) in the non-tailored feedback arm, and 297 (1 month and 6 months) in the tailored feedback arm. Baseline characteristics of the per-protocol sample were comparable to those of the ITT sample reported in section 3.2.3. The average score in emotional response to depressive symptoms was 7.0 (SD 1.9), and 48% reported to suffer from suicidal ideation at least several days within the last two weeks. Out of 820 participants interviewed with the SCID, 37% did not fulfil the criteria for major depressive disorder at baseline and were classified as false positive screens.

Rates of misdiagnosis (for all three study groups, six months; 3.5% - 4.9%), mistreatment (six months; 7.2%-8.3%), deterioration in depression severity (one month: 2% - 5.7%; six months; 4.1% - 6.8%), deterioration in emotional response (one month: 0.7% - 2.7%; six months; 1.4% - 2.9%), and deterioration in suicidal ideation (six months: 6.8% - 13.1%) were not higher in the feedback arms compared to the no feedback arm (RRs ranging between 0.46 and 1.96, with all ps ≥ 0.128; see Figure 4 for all RRs and respective 95% confidence intervals [CI]). Compared to no feedback, the rate for deterioration in suicidal ideation at one month was higher in the non-tailored feedback arm (RR = 1.92; 95% CI 1.14 to 3.24, $p = 0.014$), but not in the tailored feedback arm (RR = 1.26, 95% CI 0.25 to 1.94, $p = 0.427$), with rates of 12.3%, 8.1%, and 6.4% in the non-tailored, the tailored, and the no feedback arm at one month, respectively. All but one sensitivity analysis supported the significant effect in deterioration of suicidal ideation at one month, and there were no indications for differing effects in the subgroup of false positive screens. Based on exploratory post hoc analyses, baseline demographic and clinical characteristics of all participants deteriorated in any outcome at any time point were comparable to the total sample.

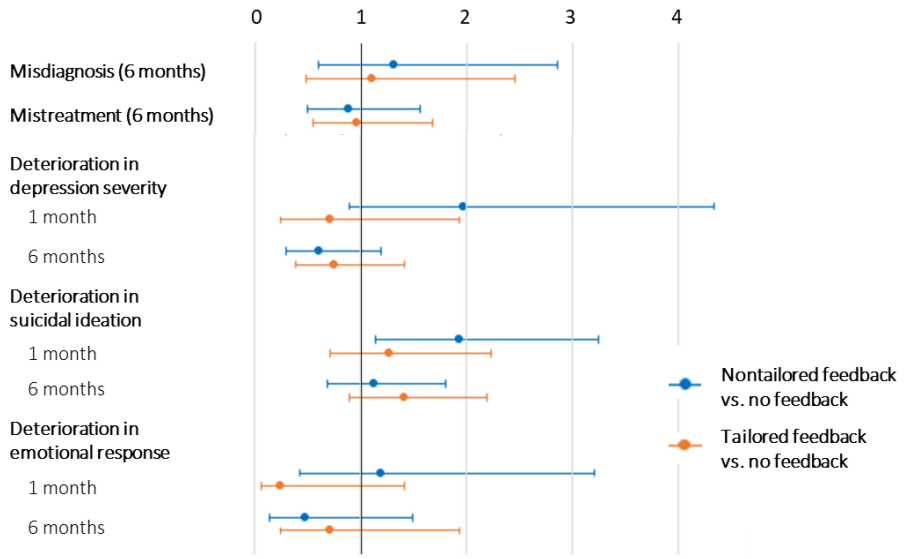


Figure 4. Relative risks (95% CIs) for all negative effects at one month and six month follow-ups in the non-tailored and tailored feedback arm as compared to no feedback (per protocol sample).

3.3.4. Limitations

First, this secondary analysis was planned post-hoc and was therefore not powered to detect differences between groups regarding selected outcomes. It cannot be ruled out that multiple testing might have led to overestimation of significance with regard to deterioration in suicidal ideation. Second, outcome selection does not consider relevant outcomes such as distress and labelling/stigma. Further, the used operationalisations of mistreatment, suicidal ideation, and emotional response to symptoms are not validated. Lastly, the findings refer to the German health care context and might differ in countries with differing health policies.

3.3.5. Conclusion

The results indicate that feedback after internet-based depression screening is not associated with healthcare-related negative effects such as misdiagnosis and mistreatment, nor with psychological negative effects such as deterioration in depression severity or in emotional response to symptoms. However, it cannot be ruled out that non-tailored feedback may increase the risk of deterioration in suicidal ideation. Against the background of the study limitations, robust prospective research on suicidal ideation in the context of internet-based depression screening is needed to inform practice as well as guidelines on (internet-based) depression screening.

3.4. Publication IV: Participants' experiences

Sikorski, F., Löwe, B., & Kohlmann, S. (2023). How adults with suspected depressive disorder experience online depression screening: A qualitative interview study. *Internet Interventions*, *34*.

3.4.1. Objective

To provide insights into the individuals' perspectives on internet-based depression screening and subsequent automated feedback, this qualitative interview study aimed to explore how adults with undiagnosed but suspected depressive disorder experience the screening process.

3.4.2. Method

This explorative qualitative interview study was conducted with a subsample of 26 participants of the DISCOVER RCT. Recruitment was conducted on an ongoing basis at the end of the 6 month follow-up interview of the RCT, with participants being informed about the aim and context of the study. Selection of participants was based on maximum variation in gender, age, study arm, and, if feasible, depression history and depression severity. Interviews were conducted via telephone (mean length = 37 min) and were informed by a semi-structured interview guide, with additional focus on what the interviewees identified as meaningful. The interviews focused on the overall experience of the screening and feedback process, rather than on the differential effects of feedback modes and screening questions only. Interviews were audiotaped, transcribed verbatim, pseudonymised, and analysed using reflexive thematic analysis (Braun & Clarke, 2006, 2019). The analysis was guided by a critical realist epistemology, aiming to uncover the underlying structures and mechanisms behind observable phenomena, while recognising knowledge as shaped by subjective interpretation. Initial coding was conducted inductively, identifying key themes and patterns in the interviews. These codes were first organised into broader themes, which were then integrated into final themes. Data collection and analysis were led by the first author (FS), who is conducting a psychodynamic psychotherapy and a PhD training, and supervised by the last author (SK), who is a cognitive-behavioural therapist and senior clinical researcher experienced with qualitative research.

3.4.3. Results

Data collection was conducted between July 2021 and August 2022. The 26 participants were balanced in terms of gender (n_{female}=15; n_{male}=11), age (ranging from 22 to 61 years), and study arm ($n_{no\ feedback}$=7; $n_{non\text{-}tailored\ feedback}$=11; $n_{tailored\ feedback}$=8).

The analysis of the interviews revealed that participants' experiences of the screening procedure can be conceptualised as a two-step process, regardless of the feedback arm. Step 1 describes the *recognition of depressive symptoms* as an initial reaction to the screening procedure. Step 2 describes a subsequent self-explorative process that encompasses up to three mutually reinforcing themes: *cognitive positioning, emotional reactions,* and *personal activation* (Figure 5).

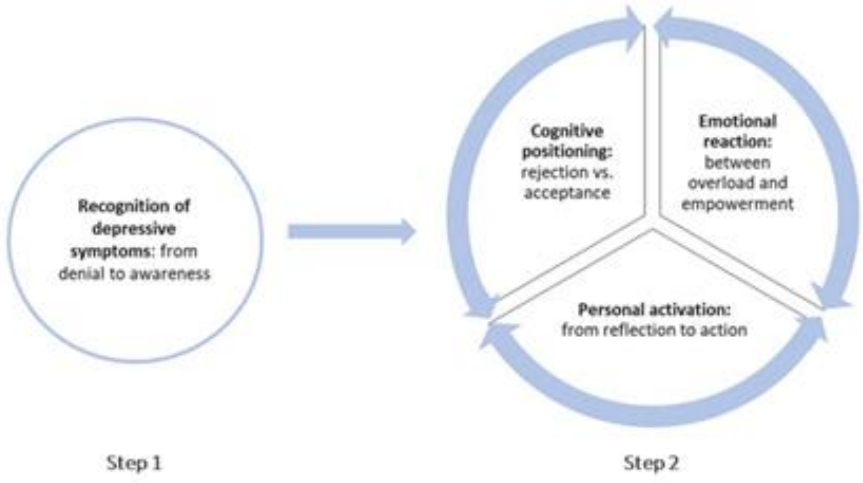


Figure 5. Participants' experiences of online depression screening alongside a two-step process.

Regarding the *recognition of depressive symptoms* (Step 1), participants described how particularly the screening questions led them to realise their depressive symptoms and symptom burden, which they had not consciously perceived, had 'denied', or had trivialised before. Regarding the subsequent self-explorative process (Step 2), some participants referred to a *cognitive positioning* towards this symptom recognition. They described how particularly the feedback triggered them to reflect on and eventually either reject or accept a depression-related self-concept. Acceptance was generally described to be 'relieving' and helpful. Often reported e*motional reactions* to both screening questions and feedback ranged from positive to challenging. Frequently, a first shock was outweighed by positive emotions such as relief,

confidence and hope. These were related to the perception of ‘not being alone’, ‘being seen’, having an explanation for the condition, or feeling validated that the condition is ‘really severe’ and not ‘imagined’ or self-caused. Challenging emotions included sadness and self-pity, partly associated with self-reproaches to not have taken the condition seriously before. Few participants further reported that specifically the screening questions triggered intense negative emotions and distressing memories from the past that required extensive coping. The last theme subsumes how participants’ reports differed in the degree of *personal activation* in response to both screening questions and feedback. Participants described a variety of actions, such as self-reflection, which enabled greater self-understanding and cognitive solution-seeking, as well as active support-seeking, that included self-management, self-care, and seeking social or professional help.

3.4.4. Limitations

Due to the self-selection of the subsample, representativeness of the results for the total RCT sample as well as the general population is limited. Further, conclusions on differential effects of the feedback modes vs no feedback is limited due to undifferentiated reports of participants and combined analysis of all three study arms. Lastly, as interviews were conducted six months after screening, participants’ memories of the screening process might have been biased by other trial-related experiences (e.g. multiple assessments and interviews) and retrospective recall effects.

3.4.5. Conclusion

Adults with undiagnosed suspected depressive disorder appear to experience online depression screening combined with feedback as a process promoting symptom recognition and subsequent self-exploration. While few participants reported intense distress associated with the screening questions, the majority described both screening questions and feedback as insightful, empowering, and activating. These findings indicate that online depression screening combined with feedback has direct subjective benefits that go beyond prompting subsequent service uptake. On the other hand, both the screening alone as well as the feedback may cause negative emotional reactions. Future research should determine to what extent online depression screening may provide a standalone form of low-threshold support for affected individuals, while taking into account the risk for potential negative effects.

4 INTEGRATION OF FINDINGS

In this section, I integrate findings of publications II to IV. Guided by the overall research question (Section 1.4) and based on qualitative findings suggesting feedback-independent effects of the depression screening, this integration addresses benefits and harms of both depression screening and automated feedback. In the following, the feedback evaluation is not differentiated by feedback mode (non-tailored vs tailored), as the study results found no evidence of differing effects based on the mode. The findings can be summarised in nine overarching themes. Benefits are described alongside *symptom reduction, quality of life, patient activation, self-acceptance and relatedness,* and *self-awareness and insight* (Table 2). Harms are described alongside *symptom deterioration and emotional burden*, *concern about symptoms*, *misinformation*, and *misallocation of healthcare* (Table 3).

Table 2. Integrated findings regarding benefits of internet-based depression screening and feedback, based on publications II and IV.

Theme	Quantitative findings	Qualitative findings (*corresponding theme*)	Meta-inferences and interpretation
Symptom reduction	At six months, **depression severity**, **anxiety severity**, and **somatic symptom severity** decreased, with **no significant differences between study arms**[1].	n/a	Quantitative findings show that **automated feedback had no impact on symptom reduction.**.
Quality of life	At six months, **quality of life** increased, with **no significant difference between study arms**[1].	n/a	Quantitative findings show that **automated feedback had no impact on quality of life..**
Patient activation	At six months, up to 17% had a new **diagnosis of depressive disorder**, and up to 29% had initiated **depression treatment**, with **no indicated difference between study arms**[1]. At six months, reported initiation of depression-related health behaviour was up to **56% for information-seeking**, up to **67% for self-management**, up to **68% for seeking social support**, and up to **49% for seeking formal help**, with **no indicated difference between study arms**[1]. In the feedback arms, 34% to 36% engaged with the feedback, and 58% to 59% downloaded the feedback[1.]	Participants reported on how **both the screening questions and the feedback** prompted self-reflection, which enabled greater self-understanding, **cognitive solution-seeking**, and **action-taking** including **self-management**, **self-care**, and **seeking social or professional help** (*theme: personal activation*)[2].	Quantitative findings demonstrate that automated feedback did not improve patient activation. Qualitative findings partly divert on that, as in some cases feedback was reported to motivate patient activation. Further, they expand on quantitative findings by indicating that often the **screening questions alone prompted patient activation** and that **activation occurred rather with regard to self-reflection and solution-seeking than to professional help-seeking**.

Table 2 (continued). Integrated findings regarding benefits of internet-based depression screening and feedback, based on publications II and IV.

Theme	Quantitative findings	Qualitative findings (*corresponding theme*)	Meta-inferences and interpretation
Self-acceptance and relatedness	n/a	Participants described how **particularly the feedback** triggered them to reflect and eventually reject or accept a depression-related self-concept, **with acceptance described to be relieving and helpful** (*theme: cognitive positioning*)[2]. Participants described that **both screening and feedback** were associated with **positive emotions such as relief, confidence and hope**. These were related to the perception of '**not being alone**', '**being seen**', **having an explanation for the condition**, or **feeling validated** that the condition is 'really severe' and not 'imagined' or self-caused (*theme: emotional reactions*)[2].	Complementarily to predefined quantitative outcomes, qualitative findings show that both **feedback and screening questions were described as having increased self-acceptance and a feeling of relatedness**, which was associated with positive emotions. (
Self-awareness and insight	n/a	Participants described how **particularly the screening questions** led them **realise their condition and their depressive symptoms** which they did not consciously perceive, denied, or trivialised before (*theme: recognition of depressive symptoms*)[2]. Participants reported on how **both the screening questions and the feedback** prompted **self-reflection, which enabled greater self-understanding, cognitive solution-seeking, and action-taking** (*theme: personal activation*)[2].	Complementarily to predefined quantitative outcomes, qualitative findings indicate that particularly the **screening questions appeared to have increased self-awareness** with regard to feelings, symptoms, and mechanisms to trivialise symptoms. Partly, participants described this self-awareness to enable insights, solution-seeking and action-taking.

Note. Meta-inferences interpret the fit between findings with regard to convergence, divergence, complementarity, or expansion. n/a=not applicable.
[1] Publication II – main analyses on efficacy. [2] Publication IV – qualitative interview study.

With regard to benefits, the integration of results reveals that quantitative and qualitative findings tend to cover different aspects of benefits: *Symptom reduction* and *quality of life* were only addressed by quantitative results, which show no significant effect for automated feedback, but indications for an improvement over time across all study arms. Qualitative findings complemented these pre-defined outcomes by reports on increased *self-awareness and insights*, *self-acceptance, and a feeling of relatedness*. Notably, qualitative findings revealed that it was not solely the feedback, but also the screening questions themselves, that influenced these outcomes. Regarding *patient activation*, quantitative results suggest that feedback did not stimulate increased depression-related health behaviour on the group level. Qualitative findings, however, indicate an effect on the individual level, as some participants indeed reported on professional help-seeking motivated by the feedback form.

Again, the qualitative findings indicate that also the screening questions promoted patient activation such as intrapsychic activation (e.g., self-reflection).

Table 3. Integrated findings regarding harms of internet-based depression screening and feedback, based on publications II-IV.

Theme	Quantitative findings	Qualitative findings	Meta-inferences and interpretation
Symptom deterioration and emotional burden	At one and six months, up to **6% and 7% deteriorated in depression severity**, with **no significant difference between feedback and no feedback**[1]. **At one month**, up to **13% deteriorated in suicidal ideation**, with the **risk for deterioration being potentially increased in the non-tailored feedback arm**. **At six month**, up to **8% deteriorated in suicidal ideation**, with **no significant difference between feedback and no feedback**[1].	At six months, less than 1% of the total sample reported negative events attributed to the trial: trial participation was **emotionally burdensome** (all study arms), associated with **distressing memories** (both feedback arms), or associated with a **feeling of helplessness** (tailored feedback)[2]. Regarding both screening questions alone and feedback, participants reported on **challenging emotions** including sadness and self-pity, which were partly associated with **self-reproaches** to not have taken the condition seriously before. Few participants also reported that the screening questions triggered **intense negative emotions** and **distressing memories from the past that required extensive coping** (*theme: emotional reactions)* [3].	Quantitative findings indicate that **automated feedback did not have a negative impact on overall symptom deterioration, but potentially on suicidal ideation**. Qualitative accounts complement on that by indicating that the screening procedure may induce **substantial emotional distress.** Contrary to quantitative findings on potential suicidal ideation, this distress is not reported in association with the feedback, but **with the screening questions alone**.
Concern about symptoms	**At one and six months**, up to **3% deteriorated in emotional response to depressive symptoms**, with **no significant difference between feedback and no feedback**[1].	Participants described that frequently, **a first shock after answering the screening questions and/or receiving the feedback was outweighed by positive emotions** such as relief, confidence and hope (*theme: emotional reactions*)[3].	Quantitative and qualitative findings agree that **automated feedback did not have an impact on concern about depressive symptoms**, which was further low across all study arms**.** Qualitative findings additionally highlight that initial concerns, if present, were mostly compensated by beneficial effects of screening questions and feedback.

Note. Meta-inferences interpret the fit between findings with regard to convergence, divergence, complementarity, or expansion. n/a=not applicable.
[1] Publication III – secondary analyses on negative effects. [2] Publication II – main analyses on efficacy.
[3] Publication IV - qualitative interview study.

Table 3 (continued). Integrated findings regarding harms of internet-based depression screening and feedback, based on publications II to IV.

Theme	Quantitative findings	Qualitative findings	Meta-inferences and interpretation
Misinformation	**37%** were screened false positive and **received false-positive feedback**[1].	Participants described how particularly the feedback triggered them to reflect and eventually **reject or accept a depression-related self-concept** (*theme: cognitive positioning*)[3].	High rates of false-positive feedback resulted in participant **misinformation**. However, expanding qualitative findings suggest that the **impact of misinformation by false-positive feedback may have been limited** as participants questioned and sometimes rejected the feedback.
Healthcare misallocation	At six months, up to **5% may have been misdiagnosed** by a health care professional as having a depression, with **no difference between feedback and no feedback**[1]. At six months, up to **8% may have been mistreated** with psychotherapy or antidepressant medication, with **no difference between feedback and no feedback**[1].	n/a	Quantitative findings indicate that **automated feedback did not have an impact on subsequent misdiagnosis or mistreatment**. Further, healthcare misallocation was low across all study arms, particularly compared with the high rate of false-positive feedback. An explanation for this discrepancy might be the postulated limited impact of false-positive feedback due to participants' questioning of the feedback (see above).

Note. Meta-inferences interpret the fit between findings with regard to convergence, divergence, complementarity, or expansion. n/a=not applicable.
[1] Publication III – secondary analyses on negative effects. [2] Publication II – main analyses on efficacy.
[3] Publication IV - qualitative interview study.

With regard to harms, findings show that despite high rates of false-positive feedback (37%) and resulting participant misinformation, automated feedback did not increase the risk for *healthcare misallocation*, *deterioration in depressive symptoms*, or *concern about symptoms*. However, it cannot be ruled out that the feedback negatively affected suicidal ideation. Although qualitative findings could not further elucidate this unexpected increase in regard to suicidal ideation, they go beyond quantitative results by revealing that solely answering the screening questions appeared to induce substantial emotional distress in some participants. Further, the qualitative finding that some participants questioned and rejected the feedback show that individuals do not necessarily believe the feedback, which can potentially explain why the negative impact of the false-positive feedback may have been limited.

5 DISCUSSION

The overall aim of this cumulative dissertation was to report and integrate findings on benefits and harms of an automated feedback after internet-based depression screening intervention within a mixed methods evaluation in adults with suspected but undiagnosed depressive disorder. Based on the rigorously designed study protocol (publication I), findings were reported in three publications investigating the intervention's efficacy (publication II), its potential negative effects (publication III), and the participants' experiences of the screening process (publication IV). The integration of findings was conducted at the reporting and interpreting level.

5.1. Critical reflection of main findings

The DISCOVER trial highlights that internet-based depression screening can potentially reach a population that is undetected but affected by averagely severe depressive symptoms, including suicidal ideation. Although the symptom severity decreased toward the end of the follow-up period, on average participants still scored above the PHQ-9 cutoff of 10 indicating the presence of a major depressive disorder, with most being undiagnosed and untreated even after six months. These results are in line with evidence on high rates of undetected depressive disorders (Beesdo-Baum et al., 2018; Trautmann & Beesdo-Baum, 2017; Vigo et al., 2020) and underline the relevance of the early detection and intervention of depressive disorders.

Contrary to trial hypotheses, the present evaluation shows that across several outcomes and irrespective of the feedback mode (non-tailored vs tailored), automated feedback following positive internet-based depression screening did not lead to substantial benefits in this target group. Compared to no feedback, providing feedback did not enhance patient activation in seeking informal or formal help. Similarly, no improvements were observed in service uptake or depression diagnosis rates, nor in symptom reduction or improvement of quality of life. However, qualitative findings complemented pre-defined outcomes and suggest that the screening process contributed to increases of the individuals' self-awareness as well as self-acceptance and a feeling of relatedness. Notably, these effects were observed in relation to both depression screening alone and the feedback form, yet it is unclear to what extent these effects can be distinctly attributed to each component. Regarding harms, our findings suggest that despite relatively high rates of misinformation of individuals due to false positive screening, the feedback of screening results did not lead to increased rates of

misdiagnosis or mistreatment. Further, the feedback intervention likely did not affect the participants' concern about their symptoms or their overall depressive symptom burden. However, it cannot be ruled out that the feedback intervention was associated with deterioration in suicidal ideation after one month. Moreover, depression screening alone contributed to emotional burden in some participants.

Our null findings regarding an effect of automated feedback on patient activation are in line with the only other trial on feedback after internet-based screening, which showed that feedback did not lead to increased uptake of mental health care when compared with a generic advice to seek help (Batterham et al., 2016). Beyond that, our findings highlight that feedback did not have substantial benefits even when providing the option to make a direct and timely appointment with a health care specialist, which is covered by the social health insurance. Interestingly, the findings contradict evidence from out research team in medical care settings, in which patients underwent depression screening in the waiting room and directly received similar, but printed feedback by study staff. In these settings, feedback has been shown to improve the patient-practitioner communication, treatment initiation (Löwe et al., 2024), and partly even depression severity (Löwe et al., 2016; Löwe et al., 2024). There are some potential explanations for this discrepancy. First, when screened in medical care settings, the individual is already 'on the spot' and automatically in contact with a health specialist. In contrast to internet-based screening, this reduces the barrier of actively seeking and realising an appointment. Second, offering patients depression screening in medical care settings might function as an explicit invitation to talk about mental problems on the part of the respective medical practitioner. This might reduce insecurities or shame to present mental problems to a health care specialist, which has been reported to be a relevant barrier to help-seeking in depression (see Doblyte & Jimenez-Mejias, 2017, for a qualitative synthesis). In contrast, when receiving feedback on the internet, the individual still needs to do the first step in approaching a health care specialist. Third, only one third of the participants engaged with the feedback (i.e. clicked on a hyperlink), and qualitative findings show that some individuals questioned and eventually rejected the validity of the feedback. It might be that in contrast to the medical care setting, a low-threshold and one-time feedback on the internet reaches less salience or is taken less serious. Lastly, the rationale underlying the feedback intervention is that it helps affected individuals recognise that they have depressive symptoms, i.e. that individuals who initially failed to connect their symptoms to a depressive disorder do so after the feedback. However, unexpectedly, in our sample 86% of participants already thought that they might currently suffer from a depressive disorder prior to screening. While this may have

complicated the statistical detection of differences between the study arms due to ceiling effects, it also suggests that, in our target group, misappraisal of symptoms may not be the primary factor contributing to delays in self-referral. In contrast, however, the qualitative study identified insights and increased awareness of symptoms as a major benefit of the screening process. Reasons for this discrepancy may lay in the self-selection of the qualitative subsample; potentially particularly those who benefitted from the screening agreed to participate in the follow-up.

The findings regarding subjectively perceived benefits such as increased self-awareness and insight, self-acceptance, and a feeling of relatedness in relation to the screening procedures expand on prior qualitative evidence. In a focus group study exploring online depression screening in young adults, participants described similar positive emotional reactions to the screening (Kruzan et al., 2022). Similarly, in studies on depression screening in primary care or postnatal settings, participants highlighted an increased awareness of symptoms and a deeper self-understanding following the depression screening procedures (Dowrick et al., 2009; Shakespeare, Blake, & Garcia, 2003; Wittkampf et al., 2008). In expansion to prior research, our findings categorised these effects, thereby highlighting self-awareness, self-acceptance and a feeling of relatedness as relevant but until now neglected patient-oriented outcomes in depression screening research. Notably, the participants in our study often reported these benefits particularly in relation with the screening questions alone. While systematic research on such mere-measurement effects in the context of depression questionnaires is missing (Preston et al., 2022), studies in other research areas support the assumption that merely asking specific questions can influence symptom perceptions or behaviours on the same topic (Godin et al., 2010; Godin, Sheeran, Conner, & Germain, 2008; Lineweaver et al., 2021).

Regarding harms, this evaluation clearly dispels the common criticism that feedback after internet-based depression screening increases misdiagnosis and mistreatment (see Danczak, 2017; Duckworth & Gilbody, 2017; Thombs et al., 2019). This does not surprise when taking into account the here reported null effects regarding service uptake. However, our evaluation does not rule out that feedback after internet-based depression screening triggers suicidal ideation. This is supported by observational evidence suggesting that particularly referrals to in-person care after internet-based screening may increase subsequent online searches for suicidal intent (Jacobson et al., 2022). In contrast, the randomised controlled trial conducted in primary care concluded that there is no indication for feedback after depression screening to increase suicidality (Löwe et al., 2024). One potential

explanation for this discrepant effect, if present, may be that in medical care settings potential emotional distress triggered by the feedback can be directly addressed, while screened individuals on the internet might be overwhelmed by dealing with emerging emotions alone. Additionally, our qualitative findings indicate that (also) the depression screening alone may trigger substantial emotional distress. Similar negative effects of depression screening have been reported in prior studies in medical care settings (Dowrick et al., 2009; Shakespeare et al., 2003; Wittkampf et al., 2008). The findings are supported by a qualitative synthesis on help-seeking in depression concluding that the recognition of symptoms as visible, real and abnormal is often associated with feelings of shame, weakness, fear, or failure, as it is perceived as a threat to one's identity (Doblyte & Jimenez-Mejias, 2017).

5.2. Strengths and limitations

This dissertation adds empirical evidence to a relatively unexplored but increasingly relevant research area. To the best of our knowledge, DISCOVER is the first research project that provides methodologically sound and empirically robust evidence on the efficacy of an automated feedback after internet-based depression screening intervention. We also conducted the first examination of related negative effects within a randomised controlled trial, and the first mixed methods evaluation of benefits and harms. The integration of findings revealed new patient-oriented outcomes of interest and novel insights on potential mere-measurement effects of depression screening alone. In all studies, we have shown great methodological thoroughness in accordance with the respective study design and the established recommendations for conducting and reporting the studies (e.g. CONSORT guidelines).

However, this mixed methods evaluation should also be considered in the context of its limitations. In addition to already described limitations in the respective publication summaries (Section 3), the first overarching limitation relates to consequences of a potential selection bias in both the RCT and the qualitative interview study. With regard to the RCT, a high proportion of participants already thought that they might currently suffer from a depressive disorder, with our actual target group, i.e. those being unaware of their depressive symptoms, being underrepresented. This may have undermined the assumed mechanism of action of the intervention, i.e. increasing the individuals' recognition that they have depressive symptoms. Regarding the qualitative interview study, the findings highlight benefits and harms that go beyond pre-defined outcomes and hypotheses, but are not generalisable to the total study sample. Specifically, they do not inform on the frequency and effect sizes of

described phenomena on the group level or whether they appeared mainly in the subgroup of participants who did not identify with a depressive disorder prior to the screening. Future research should examine these outcomes quantitatively and better match the intervention with the target group, by either addressing those unaware of their depression status or by adapting the intervention rationale to those assuming to be affected but not seeking care for other reasons. Second, against the background of reported effects of screening questions alone, it appears possible that also the mere trial participation with regular assessments and clinical contact might have had similar effects. This might have diminished intervention effects between the study arms and could contribute to explaining null findings regarding the efficacy of the feedback. Importantly, it also raises the question of whether the intervention studied should not rather have been conceptualised and evaluated as a multicomponent intervention differentiating between trial participation, survey and screening questions, and feedback of results. Unfortunately, our design precluded the ability to test the independent effects of these components. Third, the resulting integrated findings revealed that mostly, quantitative and qualitative methods examined different outcome areas. As quantitative and qualitative data collection and analyses were done separately and completed before integration of results, we missed the opportunity to use the qualitative method to explore and potentially better explain surprising phenomena such as the missing effect of feedback on depression-related outcomes or the potential effect of feedback on suicidal ideation. Last, until now a thorough evaluation of assumed mechanisms of change in our feedback intervention is lacking. Against the background of the underlying Common Sense Model of health and illness behaviour (Leventhal, Meyer, & Nerenz, 1980), examining associations between feedback and pre-post changes in illness perceptions, particularly the belief of having a depressive disorder, should be the next step to better explain and theoretically substantiate the results.

5.3. Implications

Based on this overarching evaluation and considering its limitations, a number of implications can be drawn. First, our findings underline that internet-based depression screening can reach those undiagnosed but affected, but at least in a population mostly already assuming to have a depressive disorder, a subsequent feedback intervention does not appear to have substantial benefits on depression-related outcomes. Thus, the question of how to bridge the gap between internet-based detection and initiation of depression care in this target group remains. As in internet-based treatment programmes guidance has been shown to be effective to improve patient engagement and thereby outcomes (Karyotaki et al., 2021), one approach

could be to combine a feedback after internet-based screening intervention with human interaction. For example, telephone- or chat-based counselling and guidance directly after receiving a positive feedback could be integrated. Another solution could be to minimise the barrier to follow-up depression care by directly offering effective internet-based treatment to every individual who screens positive. This pragmatic approach, however, undermines the necessity of a profound diagnosis as well as individual indications regarding personalised treatment selection (Karyotaki et al., 2021). To better understand why automated feedback appears not to have an effect on service uptake in affected individuals mainly aware of their depressive disorder, future research should focus on delays in self-referral specifically in this target group. Particularly qualitative approaches exploring how individuals process feedback of screening results directly after receiving it in addition to RCTs might be promising in this regard.

Second, until now depression screening was mainly conceptualised as aiming at improving the early detection of depressive disorders, i.e. as a 'means to another end'. In contrast, our findings regarding self-awareness and insight, self-acceptance, and the feeling of relatedness indicate that solely answering internet-based depression screening questions can have direct subjective benefits itself. This indicates that irrespective of how screening results are fed back to users, internet-based screening approaches could constitute a standalone form of low-threshold support to empower affected individuals. This finding should be considered by health care providers already offering internet-based depression screening and by internet-based information platforms and patient organisations. Future research should therefore investigate the size and generalisability of these beneficial effects. In addition, feedback after depression screening interventions should be conceptualised as multicomponent interventions and evaluated in respective designs that enable disentangling mere-measurement effects from trial participation and feedback effects. Against the background of the heterogeneity of depressive disorders (Eiko I. Fried & Nesse, 2015; Goldberg, 2011) and low content overlap among common depression scales (Eiko I Fried, 2017), future internet-based depression screening approaches should further consider selecting depression scales with regard to optimal symptom coverage rather than prognostic accuracy.

Lastly, our findings suggest that both internet-based screening as well as a subsequent feedback of results may induce harms. Further robust research on suicidal ideation and emotional distress is critical to corroborate these results. A question that follows relates to the acceptability of harms against the background of the limited benefits in internet-based depression screening. Although ethical and legal challenges of medical screening approaches

are increasingly debated in the scientific literature, until now these debates are characterised by controversial opinions and a lack of a specific legal perspective and empirical data, highlighting the lack of a consensus position in this question (see Müller et al., 2022, for a related scoping review). However, there are many websites that already publicly host internet-based depression screening interventions that are not professionally curated, potentially already representing considerable risk to vulnerable individuals. Further, in contrast to other medical software, such as for example internet-based psychotherapy programmes, medical screening is typically not subject to any level of regulation. Against this background, our findings support claims for the development of guidance on how internet-based screening should be regulated, potentially considering licensing and certification procedures, ongoing monitoring, and a legal framework (Müller et al., 2022). Until such regulations are implemented, informing service users about benefits and potential harms of internet-based depression screening would be a first step to enable users' informed decision making. These notions should be critically considered by health care providers offering internet-based depression screening and by guideline developers recommending that all adults should be screened for depression.

5.4. Conclusion

This dissertation clearly indicates that automated feedback after internet-based depression screening does not lead to substantial benefits across several depression-related outcomes including behavioural patient activation, service uptake, or symptom reduction. On the other hand, feedback can potentially lead to harms such as deterioration in suicidal ideation. As such, feedback alone does not suffice to improve early detection and treatment uptake of affected individuals. However, both screening questions and feedback demonstrate potential to empower individuals by enhancing self-awareness, self-acceptance and relatedness. As answering depression screening questions may be emotionally burdensome for affected but undiagnosed individuals, there is a need for approaches that leverage the aforementioned empowering aspects while mitigating potential harms of internet-based depression screening interventions. The findings on patient-oriented outcomes such as self-acceptance as well as harms regarding emotional burden should be corroborated by further research. In addition, research conceptualising the intervention as a multicomponent intervention could be promising for differentiating depression screening from feedback effects and allow for including new components such as human counselling. Further, practical efforts should prioritise patient safety through improved regulation and monitoring of existing

internet-based screening services. Also, health-care providers offering online depression tests should consider informing service users about benefits and harms to enable informed decision making. Altogether, this dissertation highlights the absence of a straightforward solution for addressing undetected depression while underscoring the need for future research to better understand the pathway from identification to effective management of depressive disorders.

Acknowledgment

The final stages of linguistic revision of this dissertation were partly supported by a generative language model, without contributing to the development of ideas, research findings, or intellectual content.

6 REFERENCES

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

Andersson, G. (2018). Internet interventions: Past, present and future. *Internet Interventions, 12*, 181-188. doi:https://doi.org/https://doi.org/10.1016/j.invent.2018.03.008

Barry, M. J., Nicholson, W. K., Silverstein, M., Chelmow, D., Coker, T. R., Davidson, K. W., . . . Wong, J. B. (2023). Screening for Depression and Suicide Risk in Adults: US Preventive Services Task Force Recommendation Statement. *JAMA, 329*(23), 2057-2067. doi:https://doi.org/10.1001/jama.2023.9297

Batterham, P. J., Calear, A. L., Sunderland, M., Carragher, N., & Brewer, J. L. (2016). Online screening and feedback to increase help-seeking for mental health problems: population-based randomised controlled trial. *BJPsych Open, 2*(1), 67-73. doi:https://doi.org/10.1192/bjpo.bp.115.001552

Baumeister, S. E., Schomerus, G., Schmidt, C. O., Möckel, F., van den Berg, N., Hoffmann, W., . . . Grabe, H. J. (2015). Change in depressive symptoms and mental health-related quality of life in northeast Germany between 1997-2001 and 2008-2012. *International journal of public health, 60*(1), 33-39. doi:https://doi.org/10.1007/s00038-014-0632-0

Beck, A., Hamel, C., Thuku, M., Esmaeilisaraji, L., Bennett, A., Shaver, N., . . . Stevens, A. (2022). Screening for depression among the general adult population and in women during pregnancy or the first-year postpartum: two systematic reviews to inform a guideline of the Canadian Task Force on Preventive Health Care. *Systematic Reviews, 11*(1), 176. doi:https://doi.org/10.1186/s13643-022-02022-2

Beesdo-Baum, K., Knappe, S., Einsle, F., Knothe, L., Wieder, G., Venz, J., . . . Bergmann, A. (2018). Wie häufig werden Patienten mit depressiven Störungen in der hausärztlichen Praxis erkannt? *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz, 61*(1), 52-64. doi:https://doi.org/10.1007/s00103-017-2662-2

Beesdo-Baum, K., Zaudig, M., & Wittchen, H.-U. (2019). *SCID-5-CV Strukturiertes Klinisches Interview für DSM-5-Störungen–Klinische Version: Deutsche Bearbeitung des Structured Clinical Interview for DSM-5 Disorders–Clinician Version von Michael B. First, Janet BW Williams, Rhonda S. Karg, Robert L. Spitzer*: Hogrefe.

Berger, M., van Calker, D., Brakemeier, E. L., et al. (2019). Affektive Störungen. In M. Berger (Ed.), *Psychische Erkrankungen: Klinik und Therapie in Zusammenarbeit mit der Cochrane Deutschland Stiftung* (6th ed., pp. 363–435). Elsevier.

BinDhim, N. F., Alanazi, E. M., Aljadhey, H., Basyouni, M. H., Kowalski, S. R., Pont, L. G., . . . Alhawassi, T. M. (2016). Does a Mobile Phone Depression-Screening App Motivate Mobile Phone Users With High Depressive Symptoms to Seek a Health Care Professional's Help? *Journal of Medical Internet Research, 18*(6), e156. doi:https://doi.org/10.2196/jmir.5726

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101. doi:https://doi.org/10.1191/1478088706qp063oa

Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health, 11*(4), 589-597. doi:https://doi.org/10.1080/2159676X.2019.1628806

Bundesärztekammer (BÄK), Kassenärztliche Bundesvereinigung (KBV), Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF). (2022). *Nationale VersorgungsLeitlinie Unipolare Depression – Langfassung, Version 3.0*. [cited: 2024-01-05]. doi:https://doi.org/10.6101/AZQ/000493. www.leitlinien.de/depression.

Chesney, E., Goodwin, G. M., & Fazel, S. (2014). Risks of all-cause and suicide mortality in mental disorders: a meta-review. *World Psychiatry, 13*(2), 153-160. doi:https://doi.org/10.1002/wps.20128

Choi, I., Milne, D. N., Deady, M., Calvo, R. A., Harvey, S. B., & Glozier, N. (2018). Impact of Mental Health Screening on Promoting Immediate Online Help-Seeking: Randomized Trial Comparing Normative Versus Humor-Driven Feedback. *JMIR Mental Health, 5*(2), e26. doi:https://doi.org/10.2196/mental.9480

Cipriani, A., Furukawa, T. A., Salanti, G., Chaimani, A., Atkinson, L. Z., Ogawa, Y., . . . Geddes, J. R. (2018). Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. *Lancet, 391*(10128), 1357-1366. doi:https://doi.org/10.1016/s0140-6736(17)32802-7

Cuijpers, P., Miguel, C., Harrer, M., Plessen, C. Y., Ciharova, M., Papola, D., . . . Karyotaki, E. (2023). Psychological treatment of depression: A systematic overview of a 'Meta-Analytic Research Domain'. *Journal of Affective Disorders, 335*, 141-151. doi:https://doi.org/https://doi.org/10.1016/j.jad.2023.05.011

Danczak, A. (2017). Online screening test for depression is inappropriate. *BMJ, 359*, j4736. doi:https://doi.org/10.1136/bmj.j4736

Daniela, R. J., Liliane, Z., Susan, G., Yoon, L., Joel, J. G., Steven, A. J., . . . Sunita, V. (2023). CONSORT Harms 2022 statement, explanation, and elaboration: updated guideline for the reporting of harms in randomised trials. *BMJ, 381*, e073725. doi:https://doi.org/10.1136/bmj-2022-073725

Dawadi, S., Shrestha, S., & Giri, R. A. (2021). Mixed-Methods Research: A Discussion on its Types, Challenges, and Criticisms. *Journal of Practical Studies in Education, 2*(2), 25-36. doi:https://doi.org/10.46809/jpse.v2i2.20

Doblyte, S., & Jimenez-Mejias, E. (2017). Understanding Help-Seeking Behavior in Depression: A Qualitative Synthesis of Patients' Experiences. *Qualitative Health Research, 27*(1), 100-113. doi:https://doi.org/10.1177/1049732316681282

Dobrow, M. J., Hagens, V., Chafe, R., Sullivan, T., & Rabeneck, L. (2018). Consolidated principles for screening based on a systematic review and consensus process. *Canadian Medical Association Journal, 190*(14), E422-e429. doi:https://doi.org/10.1503/Canadian Medical Association Journal.171154

Dowrick, C., Leydon, G. M., McBride, A., Howe, A., Burgess, H., Clarke, P., . . . Kendrick, T. (2009). Patients' and doctors' views on depression severity questionnaires incentivised in UK quality and outcomes framework: qualitative study. *British Medical Journal, 338*, b663. doi:https://doi.org/10.1136/bmj.b663

Duckworth, K., & Gilbody, S. (2017). Should Google offer an online screening test for depression? *BMJ, 358*, j4144. doi:https://doi.org/10.1136/bmj.j4144

Eichenberg, C., Wolters, C., & Brähler, E. (2013). The Internet as a Mental Health Advisor in Germany-Results of a National Survey. *PloS One, 8*(11), e79206. doi:https://doi.org/ARTN e7920610.1371/journal.pone.0079206

Eysenbach, G. (2013). CONSORT-EHEALTH: implementation of a checklist for authors and editors to improve reporting of web-based and mobile randomized controlled trials. *Studies in Health Technology and Informatics, 192*, 657-661.

Fetters, M. D., Curry, L. A., & Creswell, J. W. (2013). Achieving integration in mixed methods designs-principles and practices. *Health Service Research, 48*(6 Pt 2), 2134-2156. doi:https://doi.org/10.1111/1475-6773.12117

Fried, E. I. (2017). The 52 symptoms of major depression: Lack of content overlap among seven common depression scales. *Journal of Affective Disorders, 208*, 191-197.

Fried, E. I., & Nesse, R. M. (2015). Depression is not a consistent syndrome: An investigation of unique symptom patterns in the STAR*D study. *Journal of Affective Disorders, 172*, 96-102. doi:https://doi.org/https://doi.org/10.1016/j.jad.2014.10.010

GBD 2019 Mental Disorders Collaborators. (2022). Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990–2019: A systematic analysis for the Global Burden of Disease Study 2019. *The Lancet Psychiatry, 9*(2), 137-150.

Gierk, B., Kohlmann, S., Kroenke, K., Spangenberg, L., Zenger, M., Brahler, E., & Lowe, B. (2014). The somatic symptom scale-8 (SSS-8): a brief measure of somatic symptom burden. *JAMA Internal Medicine, 174*(3), 399-407. doi:https://doi.org/10.1001/jamainternmed.2013.12179

Godin, G., Sheeran, P., Conner, M., Delage, G., Germain, M., Bélanger-Gravel, A., & Naccache, H. (2010). Which survey questions change behavior? Randomized controlled trial of mere measurement interventions. *Health Psychology, 29*(6), 636-644. doi:https://doi.org/10.1037/a0021131

Godin, G., Sheeran, P., Conner, M., & Germain, M. (2008). Asking questions changes behavior: mere measurement effects on frequency of blood donation. *Health Psychology, 27*(2), 179-184. doi:https://doi.org/10.1037/0278-6133.27.2.179

Gold, S. M., Köhler-Forsberg, O., Moss-Morris, R., Mehnert, A., Miranda, J. J., Bullinger, M., . . . Otte, C. (2020). Comorbid depression in medical diseases. *Nature reviews Disease primers, 6*(1), 69. doi:https://doi.org/10.1038/s41572-020-0200-2

Goldberg, D. (2011). The heterogeneity of "major depression". *World Psychiatry, 10*(3), 226-228. doi:https://doi.org/10.1002/j.2051-5545.2011.tb00061.x

Gunther, O. H., Roick, C., Angermeyer, M. C., & Konig, H. H. (2008). The responsiveness of EQ-5D utility scores in patients with depression: A comparison with instruments measuring quality of life, psychopathology and social functioning. *Journal of Affective Disorders, 105*(1-3), 81-91. doi:https://doi.org/10.1016/j.jad.2007.04.018

Heinz, I., Baldofski, S., Beesdo-Baum, K., Knappe, S., Kohls, E., & Rummel-Kluge, C. (2021). “Doctor, my back hurts and I cannot sleep.” Depression in primary care patients: Reasons for consultation and perceived depression stigma. *PLoS One, 16*(3), e0248069. doi:https://doi.org/10.1371/journal.pone.0248069

Herrman, H., Patel, V., Kieling, C., Berk, M., Buchweitz, C., Cuijpers, P., . . . Wolpert, M. (2022). Time for united action on depression: a Lancet–World Psychiatric Association Commission. *The Lancet, 399*(10328), 957-1022. doi:https://doi.org/https://doi.org/10.1016/S0140-6736(21)02141-3

Jacobi, F., Hofler, M., Siegert, J., Mack, S., Gerschler, A., Scholl, L., . . . Wittchen, H. U. (2014). Twelve-month prevalence, comorbidity and correlates of mental disorders in Germany: the Mental Health Module of the German Health Interview and Examination Survey for Adults (DEGS1-MH). *International Journal of Methods in Psychiatric Research, 23*(3), 304-319. doi:https://doi.org/10.1002/mpr.1439

Jacobson, N. C., Yom-Tov, E., Lekkas, D., Heinz, M., Liu, L., & Barr, P. J. (2022). Impact of online mental health screening tools on help-seeking, care receipt, and suicidal ideation and suicidal intent: Evidence from internet search behavior in a large U.S. cohort. *Journal of Psychiatric Research, 145*, 276-283. doi:https://doi.org/10.1016/j.jpsychires.2020.11.010

Karyotaki, E., Efthimiou, O., Miguel, C., Maas genannt Bermpohl, F., Furukawa, T. A., Cuijpers, P., & Individual Patient Data Meta-Analyses for Depression, C. (2021). Internet-Based Cognitive Behavioral Therapy for Depression: A Systematic Review and Individual Patient Data Network Meta-analysis. *JAMA Psychiatry, 78*(4), 361-371. doi:https://doi.org/10.1001/jamapsychiatry.2020.4364

König, H., König, H. H., & Konnopka, A. (2020). The excess costs of depression: a systematic review and meta-analysis. *Epidemiology and Psychiatric Sciences, 29*, e30. doi:https://doi.org/10.1017/S2045796019000180

Kratz, S., Härter, M., Bermejo, I., Berger, M., Schneider, F., & Gaebel, W. (2003). [Reason for encounter and diagnosis of depression in patients in general practice]. *Zeitschrift fur arztliche Fortbildung und Qualitatssicherung, 97 Suppl 4*, 50-56.

Kraus, C., Kadriu, B., Lanzenberger, R., Zarate, C. A., Jr., & Kasper, S. (2019). Prognosis and improved outcomes in major depression: a review. *Translational Psychiatry, 9*(1), 127. doi:https://doi.org/10.1038/s41398-019-0460-3

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine, 16*(9), 606-613. doi:https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Kruzan, K. P., Meyerhoff, J., Nguyen, T., Mohr, D. C., Reddy, M., & Kornfield, R. (2022). "I Wanted to See How Bad it Was": Online Self-screening as a Critical Transition Point Among Young Adults with Common Mental Health Conditions. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2022*. doi:https://doi.org/10.1145/3491102.3501976

Leventhal, H., Meyer, D., & Nerenz, D. (1980). The common sense model of illness danger. In S. Rachman (Ed.), *Medical psychology: Contributions to medical psychology* (pp. 7-30).

Levitt, H. M., Bamberg, M., Creswell, J. W., Frost, D. M., Josselson, R., & Suarez-Orozco, C. (2018). Journal article reporting standards for qualitative primary, qualitative meta-analytic, and mixed methods research in psychology: The APA Publications and Communications Board task force report. *American Psychologist, 73*(1), 26-46. doi:https://doi.org/10.1037/amp0000151

Lim, K. L., Jacobs, P., Ohinmaa, A., Schopflocher, D., & Dewa, C. S. (2008). A new population-based measure of the economic burden of mental illness in Canada. *Chronic Diseases in Canada, 28*(3), 92-98.

Lineweaver, T. T., Kercood, S., Gabor, A. J., Cervantes, J., Laine, A., & Baker, E. (2021). The effect of medication and question wording on self-reported symptoms and their accuracy in young adults with attention-deficit/hyperactivity disorder. *British Journal of Clinical Psychology, 60*(2), 252-269. doi:https://doi.org/10.1111/bjc.12276

Löwe, B., Blankenberg, S., Wegscheider, K., König, H.-H., Walter, D., Murray, A. M., . . . Kohlmann, S. (2016). Depression screening with patient-targeted feedback in cardiology: DEPSCREEN-INFO randomised clinical trial. *British Journal of Psychiatry*, 1-8.

Lowe, B., Decker, O., Muller, S., Brahler, E., Schellberg, D., Herzog, W., & Herzberg, P. Y. (2008). Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. *Medical Care, 46*(3), 266-274. doi:https://doi.org/10.1097/MLR.0b013e318160d093

Löwe, B., Scherer, M., Braunschneider, L.-E., Marx, G., Eisele, M., Mallon, T., . . . Kohlmann, S. (2024). Clinical effectiveness of patient-targeted feedback following depression screening in general practice (GET.FEEDBACK.GP): an investigator-initiated, prospective, multicentre, three-arm, observer-blinded, randomised controlled trial in Germany. *The Lancet Psychiatry, 11*(4), 262-273. doi:https://doi.org/https://doi.org/10.1016/S2215-0366(24)00035-X

Martin, R., Rothrock, N., Leventhal, H., & Leventhal, E. (2003). Common Sense Models of Illness: Implications for Symptom Perception and Health-Related Behaviors. In *Social Psychological Foundations of Health and Illness* (pp. 199-225).

Monroe, S. M., & Harkness, K. L. (2011). Recurrence in major depression: A conceptual analysis. *Psychological Review, 118*(4), 655-674. doi:https://doi.org/10.1037/a0025190

Montgomery, P., Grant, S., Mayo-Wilson, E., Macdonald, G., Michie, S., Hopewell, S., . . . Group, o. b. o. t. C.-S. (2018). Reporting randomised trials of social and psychological interventions: the CONSORT-SPI 2018 Extension. *Trials, 19*(1), 407. doi:https://doi.org/10.1186/s13063-018-2733-1

Müller, R., Klemmt, M., Ehni, H.-J., Henking, T., Kuhnmünch, A., Preiser, C., . . . Ranisch, R. (2022). Ethical, legal, and social aspects of symptom checker applications: a scoping review. *Medicine, Health Care and Philosophy, 25*(4), 737-755. doi:https://doi.org/10.1007/s11019-022-10114-y

Murphy, J. M., Laird, N. M., Monson, R. R., Sobol, A. M., & Leighton, A. H. (2000). A 40-Year Perspective on the Prevalence of Depression: The Stirling County Study. *Archives of General Psychiatry, 57*(3), 209-215. doi:https://doi.org/10.1001/archpsyc.57.3.209

Negeri, Z. F., Levis, B., Sun, Y., He, C., Krishnan, A., Wu, Y., . . . Thombs, B. D. (2021). Accuracy of the Patient Health Questionnaire-9 for screening to detect major depression: updated systematic review and individual participant data meta-analysis. *BMJ, 375*, n2183. doi:https://doi.org/10.1136/bmj.n2183

O'Connor, E. A., Perdue, L. A., Coppola, E. L., Henninger, M. L., Thomas, R. G., & Gaynes, B. N. (2023). Depression and Suicide Risk Screening: Updated Evidence Report and Systematic Review for the US Preventive Services Task Force. *JAMA, 329*(23), 2068-2085. doi:https://doi.org/10.1001/jama.2023.7787

Preston, L., Valentin, R., Lisa, O., Alize, R., Linetta, K., Anouk, H., & Tanja, S. (2022). Systematic literature review examining the mere-measurement effect of patient reported measures: the interaction of emotional valence and frequency of exposure as an independent factor in patient change. *medRxiv*, 2022.2003.2029.22273094. doi:https://doi.org/10.1101/2022.03.29.22273094

Ryan, A., & Wilson, S. (2008). Internet healthcare: do self-diagnosis sites do more harm than good? *Expert Opinion on Drug Safety, 7*(3), 227-229. doi:https://doi.org/10.1517/14740338.7.3.227

Schneider, F., Erhart, M., Hewer, W., Loeffler, L. A., & Jacobi, F. (2019). Mortality and Medical Comorbidity in the Severely Mentally Ill. *Dtsches Arzteblatt International, 116*(23-24), 405-411. doi:https://doi.org/10.3238/arztebl.2019.0405

Schreiter, S., Mascarell-Maricic, L., Rakitzis, O., Volkmann, C., Kaminski, J., & Daniels, M. A. (2023). Digital Health Applications in the Area of Mental Health: A Scoping Review. *Deutsches Ärzteblatt International, 120*(47), 797.

Shakespeare, J., Blake, F., & Garcia, J. (2003). A qualitative study of the acceptability of routine screening of postnatal women using the Edinburgh Postnatal Depression Scale. *Br J General Practitioner, 53*(493), 614-619.

Steffen, A., Nübel, J., Jacobi, F., Bätzing, J., & Holstiege, J. (2020). Mental and somatic comorbidity of depression: a comprehensive cross-sectional analysis of 202 diagnosis groups using German nationwide ambulatory claims data. *BMC Psychiatry, 20*(1), 142. doi:https://doi.org/10.1186/s12888-020-02546-8

Streit, F., Zillich, L., Frank, J., Kleineidam, L., Wagner, M., Baune, B. T., . . . Berger, K. (2023). Lifetime and current depression in the German National Cohort (NAKO). *The World Journal of Biological Psychiatry, 24*(10), 865-880. doi:https://doi.org/10.1080/15622975.2021.2014152

Thombs, B., Turner, K. A., & Shrier, I. (2019). Defining and Evaluating Overdiagnosis in Mental Health: A Meta-Research Review. *Psychotherapy and Psychosomatics, 88*(4), 193-202. doi:https://doi.org/10.1159/000501647

Thombs, B. D., Markham, S., Rice, D. B., & Ziegelstein, R. C. (2021). Does depression screening in primary care improve mental health outcomes? *BMJ, 374*, n1661. doi:https://doi.org/10.1136/bmj.n1661

Tong, A., Sainsbury, P., & Craig, J. (2007). Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care, 19*(6), 349-357. doi:https://doi.org/10.1093/intqhc/mzm042

Trautmann, S., & Beesdo-Baum, K. (2017). The treatment of depression in primary care: a cross-sectional epidemiological study. *Deutsches Ärzteblatt International, 114*(43), 721. doi:https://doi.org/10.3238/arztebl.2017.0721

Vaidyanathan, U., Sun, Y., Shekel, T., Chou, K., Galea, S., Gabrilovich, E., & Wellenius, G. A. (2022). An evaluation of Internet searches as a marker of trends in population mental health in the US. *Scientific Reports, 12*(1), 8946. doi:https://doi.org/10.1038/s41598-022-12952-5

Vigo, D., Haro, J. M., Hwang, I., Aguilar-Gaxiola, S., Alonso, J., Borges, G., . . . Kessler, R. C. (2020). Toward measuring effective treatment coverage: critical bottlenecks in quality- and user-adjusted coverage for major depressive disorder. *Psychological Medicine*, 1-11. doi:https://doi.org/10.1017/s0033291720003797

Wittkampf, K. A., van Zwieten, M., Smits, F. T., Schene, A. H., Huyser, J., & van Weert, H. C. (2008). Patients' view on screening for depression in general practice. *Family Practitioner, 25*(6), 438-444. doi:https://doi.org/10.1093/fampra/cmn057

World Health Organization. (2015). *mhGAP Humanitarian Intervention Guide (mhGAPHIG): Clinical management of mental, neurological, and substance use conditions in humanitarian emergencies*. World Health Organization.

World Health Organization. (2019). *International statistical classification of diseases and related health problems* (11th ed.).

World Health Organization. (2022). *World mental health report: Transforming mental health for all*. World Health Organization.

Yom-Tov, E., Lekkas, D., Heinz, M. V., Nguyen, T., Barr, P. J., & Jacobson, N. C. (2023). Digitally filling the access gap in mental health care: An investigation of the association between rurality and online engagement with validated self-report screens across the United States. *Journal of Psychiatry Research, 157*, 112-118. doi:https://doi.org/10.1016/j.jpsychires.2022.11.024

7 PUBLICATIONS

7.1. Publication I

The efficacy of automated feedback after internet-based depression screening: Study protocol of the German, three-armed, randomised controlled trial DISCOVER

Sikorski, F., König, H.-H., Wegscheider, K., Zapf, A., Löwe, B., & Kohlmann, S.

Status: Published in Internet Interventions online, July 21, 2021.

Internet Interventions 25 (2021) 100435

Contents lists available at ScienceDirect

Internet Interventions

journal homepage: www.elsevier.com/locate/invent

The efficacy of automated feedback after internet-based depression screening: Study protocol of the German, three-armed, randomised controlled trial DISCOVER

Franziska Sikorski [a,*], Hans-Helmut König [b], Karl Wegscheider [c], Antonia Zapf [c], Bernd Löwe [a], Sebastian Kohlmann [a]

[a] Department of Psychosomatic Medicine and Psychotherapy, University Medical Centre Hamburg-Eppendorf, Hamburg, Germany
[b] Department of Health Economics and Health Services Research, University Medical Centre Hamburg-Eppendorf, Hamburg, Germany
[c] Department of Medical Biometry and Epidemiology, University Medical Centre Hamburg-Eppendorf, Hamburg, Germany

ARTICLE INFO

Keywords:
Depression screening
Early detection
Tailored feedback
Patient engagement
Internet-based intervention
Randomised controlled trial protocol

ABSTRACT

Background: Depression is one of the most disabling disorders worldwide, yet it often remains undetected. One promising approach to address both early detection and disease burden is depression screening followed by direct feedback to patients. Evidence suggests that individuals often seek information regarding mental health on the internet. Thus, internet-based screening with automated feedback has great potential to address individuals with undetected depression.

Objectives: To determine whether automated feedback after internet-based depression screening reduces depression severity as compared to no feedback.

Methods: The internet-based, observer-blinded DISCOVER RCT aims to recruit a total of 1074 individuals. Participants will be screened for depression using the Patient Health Questionnaire (PHQ-9). In case of a positive screening result (PHQ-9 $\geq$ 10), participants with undetected depression will be randomised into one of three balanced study arms to receive either (a) no feedback (control arm), (b) standard feedback, or (c) tailored feedback on their screening result. The tailored feedback version will be adapted to participants' characteristics, i.e. symptom profile, preferences, and demographic characteristics. The primary hypothesis is that feedback reduces depression severity six months after screening compared to no feedback. The secondary hypothesis is that tailored feedback is more efficacious compared to standard feedback. Further outcomes are depression care, help-seeking behaviour, health-related quality of life, anxiety, somatic symptom severity, intervention acceptance, illness beliefs, adverse events, and a health economic evaluation. Follow-ups will be conducted one month and six months after screening by self-report questionnaires and clinical interviews. According to a statistical analysis plan, the primary outcome will be analysed on an intention-to-treat basis applying multilevel modelling.

Discussion: The results of the DISCOVER RCT will inform about how automated feedback after internet-based screening could improve early detection and resolution of depression. Ways of dissemination and how the trial can contribute to an understanding of help-seeking behaviour processes will be discussed. If the results show that automated feedback after internet-based depression screening can reduce depression severity, the intervention could be easily implemented and might substantially reduce the disease burden of individuals with undetected depression.

Ethical approval: The study is approved by the Ethics Committee of the Hamburg Medical Association.

Trial registration: The trial was registered at ClinicalTrials.gov in November 2020 (identifier: NCT04633096).

* Corresponding author at: Department of Psychosomatic Medicine and Psychotherapy, University Medical Centre Hamburg-Eppendorf, Martinistr. 52, 20246 Hamburg, Germany.
E-mail addresses: f.sikorski@uke.de (F. Sikorski), s.kohlmann@uke.de (S. Kohlmann).

https://doi.org/10.1016/j.invent.2021.100435
Received 24 March 2021; Received in revised form 16 July 2021; Accepted 16 July 2021
Available online 21 July 2021

1. Background

Major depression is one of the most disabling disorders worldwide and affects one out of ten individuals over their lifetime (Busch et al., 2013; Vos et al., 2020). Untreated depression leads to rising healthcare costs, has an increased likelihood of a chronic course and treatment resistance and, most importantly, results in an increased disease burden (Chisholm et al., 2016; Fichter et al., 2010; Ghio et al., 2014). Nevertheless, depression often remains undetected: in primary care, for example, it is estimated that only 50% of depressed patients are correctly diagnosed as such (Mitchell et al., 2009; Trautmann and Beesdo-Baum, 2017). One promising approach to address early detection of depression is widely accessible depression screening.

Standardised depression screening alone, however, appears to be insufficient to alter disease burden (Gilbody et al., 2008; Thombs et al., 2014). A worthwhile approach to increase the efficacy of depression screening is to enhance patient engagement by feedback provided directly to the individual. In line with self-regulation theories of health behaviour (e.g. Leventhal et al., 2003), feedback allows individuals to recognise that they suffer from depression and motivates individuals to actively engage in functional health behaviour such as help-seeking and depression care. In turn, this should reduce depression severity in the long run. Indeed, the results of our preceding DEPSCREEN-INFO RCT indicate that a feedback intervention - including the screening result as well as recommendations on further diagnostic consultation and help-seeking – can increase patients' engagement in seeking information on depression and, most importantly, reduce depression severity after six months in patients with coronary heart disease (Löwe et al., 2016).

To expand the evidence on feedback after depression screening to the primary care setting, we currently run the multicentre RCT GET.FEEDBACK.GP (Kohlmann et al., 2020). Yet, barriers such as fear of stigmatisation or the desire to handle the problem on one's own often deter professional help-seeking in depression (Boerema et al., 2016; Schomerus and Angermeyer, 2008). Whereas individuals with stigmatised symptoms may be reluctant to present to a health professional, however, the internet has increasingly become a source for individuals with elevated depression severity to actively seek mental health information (Berger et al., 2005). In Germany, for example, one of four individuals would consider seeking help for mental health online (Eichenberg et al., 2013). Conducting the feedback intervention as an internet-based intervention, therefore, appears to have a great potential to reach a large population of affected individuals outside of the medical system.

In other domains such as prevention and intervention of mental disorders, internet-based interventions have already been shown to be effective (e.g. Ebert et al., 2017; Karyotaki et al., 2017; Richards and Richardson, 2012). Additionally, they can bring the benefits of fostering anonymity, of being cost-effective, and of being scalable, thus allowing for large populations to be reached (Andersson, 2016; Andersson and Titov, 2014; Ebert et al., 2017). Notably, the internet-based format also offers the possibility to individually tailor the feedback according to individuals' characteristics (Andersson and Titov, 2014). This is promising, as compared to standard health messages, tailored messages are more frequently read, better remembered and perceived as more relevant (Ryan et al., 2001). Regarding depression, tailored health messages motivate patients to engage in depression care and can help to reduce depression severity (Levesque et al., 2011; Shah et al., 2014). Tailored feedback after depression screening offers the opportunity to match depression-related information to individuals' characteristics with the aim to make it more salient. Accordingly, tailored feedback has the potential to enhance the effect on patient engagement and depression severity compared to standardised feedback.

Here, we describe the three-armed DISCOVER RCT to address early detection and resolution of depression by testing the efficacy of automated feedback after internet-based depression screening, as compared to no feedback. In addition, we will compare the efficacy of a standardised version of the feedback with a version that is tailored to participants' symptom profiles, preferences, and sociodemographic characteristics. The primary outcome will be depression severity six months after internet-based screening. To allow for a comprehensive evaluation, further secondary outcomes and process variables will be examined.

1.1. Trial hypotheses

The primary hypothesis is that depression severity six months after screening is lower in each of the two feedback study arms (STANDARD FEEDBACK and TAILORED FEEDBACK) as compared to the NO FEEDBACK study arm. As we assume that tailored feedback can maximise the efficacy of standardised feedback, the secondary hypothesis is that depression severity six months after screening is lower in the TAILORED FEEDBACK arm as compared to the STANDARD FEEDBACK arm.

2. Methods

2.1. Design

The DISCOVER trial is designed as an internet-based, observer-blinded, randomised controlled clinical trial with three parallel groups, which is conducted nationwide in Germany. After undergoing an online depression screening with the Patient Health Questionnaire-9 (PHQ-9; Kroenke et al., 2001; Löwe et al., 2004a,b), participants with suspected depressive disorder (PHQ-9 $\geq$ 10 points) will be randomised into one of three balanced study arms: (a) NO FEEDBACK, (b) STANDARD FEEDBACK, or (c) TAILORED FEEDBACK on their screening result. Assessments will be conducted online and via telephone and will be scheduled at baseline (before randomisation: T0; 2 days after randomisation: T1), at 1-month (T2), and at 6-months follow-up (T3). The primary objective of the trial is to show superiority of both feedback arms compared to the control arm regarding depression severity 6 months after screening.

The trial (protocol) will be conducted and reported according to adequate CONSORT 2010 extensions and the CONSORT E-HEALTH statement (Boutron et al., 2017; Eysenbach, and Group, 2011; Moher et al., 2010; Montgomery et al., 2018; Schulz et al., 2010), as well as the SPIRIT 2013 statement (Chan et al., 2013).

2.2. Inclusion and exclusion criteria

Eligibility criteria will be assessed within a self-report online survey at T0. Participants will be required to (a) be aged 18 years or above, (b) have sufficient German language proficiency, (c) show an indication for at least moderate depression (PHQ-9 $\geq$ 10 points), (d) provide contact details, (e) have internet access, (f) have sufficient computer/internet literacy and (f) be willing to give informed consent. Participants will be excluded (a) if they were diagnosed with depression within the past 12 months or (b) if they currently are or were receiving depression treatment within the past 12 months.

2.3. Recruitment and procedure

The trial will be publicly promoted as a study ‘on stress and psychological well-being’. Study participants will be recruited from the general population through traditional and social media campaigns (e.g. advertisement on related websites/newsletters and Google, posts on Facebook, Instagram and Twitter) and through print advertisement in public areas of several German cities (e.g. flyers, posters). To reach a sample that strives for representativeness of the German population with respect to age and gender, a marketing company will further advertise the study via a population wide online access survey panel. Recruitment success and sample characteristics (i.e. age, gender) will be monitored on an ongoing basis and strategies will be adapted, if necessary. Recruitment has started in January 2021 and is planned to run for 12 months.

All recruitment ways will lead to the open access study website (https://www.discover-studie.de), which is designed in responsive design to ensure optimal usability for all types of devices (e.g. mobile devices, tablets). The website contains detailed information on the study, data safety procedures, the study team, and contact information. Interested applicants will be asked to provide online informed consent and thereafter to complete the T0 assessment. All participants indicating an elevated suicide risk (PHQ-9 suicide item ≥2 points) will be shown a screen with urgent advice to seek help and relevant information on available help services (e.g. general practitioner, local psychiatric emergency units, and the national emergency number). After having completed the survey, all eligible participants will be randomised and will be directly provided with feedback on their depression screening result (STANDARD and TAILORED FEEDBACK) or a 'thank you'-note (NO FEEDBACK). They will be contacted and reminded via email on the online follow-up assessments (T1–T3) and via telephone for supplemental clinical interviews (T1 and T3). Whereas the T0 assessment will not be financially rewarded, for each complete follow-up assessment participants immediately receive a compensation of five euro as a voucher (i.e. 3 × five euro vouchers in total). Fig. 1 provides a detailed overview of the study flow.

All procedures involved in the study are consistent with generally accepted standards of ethical practice such as the Declaration of Helsinki and have been approved by the Ethics Committee of the Hamburg Medical Association in July 2019 (reference number: PV7039). The trial was registered at ClinicalTrials.gov in November 2020 (identifier: NCT04633096).

2.4. Randomisation and blinding

Randomisation will be based on a computer-generated randomisation sequence (1:1:1 allocation ratio), which was conducted by an independent researcher of the Department of Medical Biometry and Epidemiology and is not accessible to any other study team member. The sequence consists of permuted blocks of randomly arranged sizes (6, 9, and 12) and is stratified by baseline depression severity (moderate: PHQ-9 ≥ 10–14 points; severe: PHQ-9 ≥ 15 points) to guarantee equity of sample sizes across study arms and severity levels. Allocation will be performed by a computerised system, ensuring allocation concealment. Individuals who participate multiple times will be automatically allocated to the same study arm as before. This process is ensured by a privacy-preserving record linkage service which identifies double entries based on personal data and the IP address (Mainzelliste; Rohde et al., 2021).

Participants will know their allocation due to the nature of the intervention but will be kept unaware of trial hypotheses to minimise expectancy bias. The research staff assessing outcomes in the telephone interviews will be blind to the allocation at any time. Steps to control for blindness include the following: after every interview, assessors are (a) instructed to document if participants have disclosed their randomisation status and (b) asked to guess the study arm. After study closure, this guess will be compared with the actual status and Cohen's kappa will be computed to identify whether hit rates differ from what can be expected from chance.

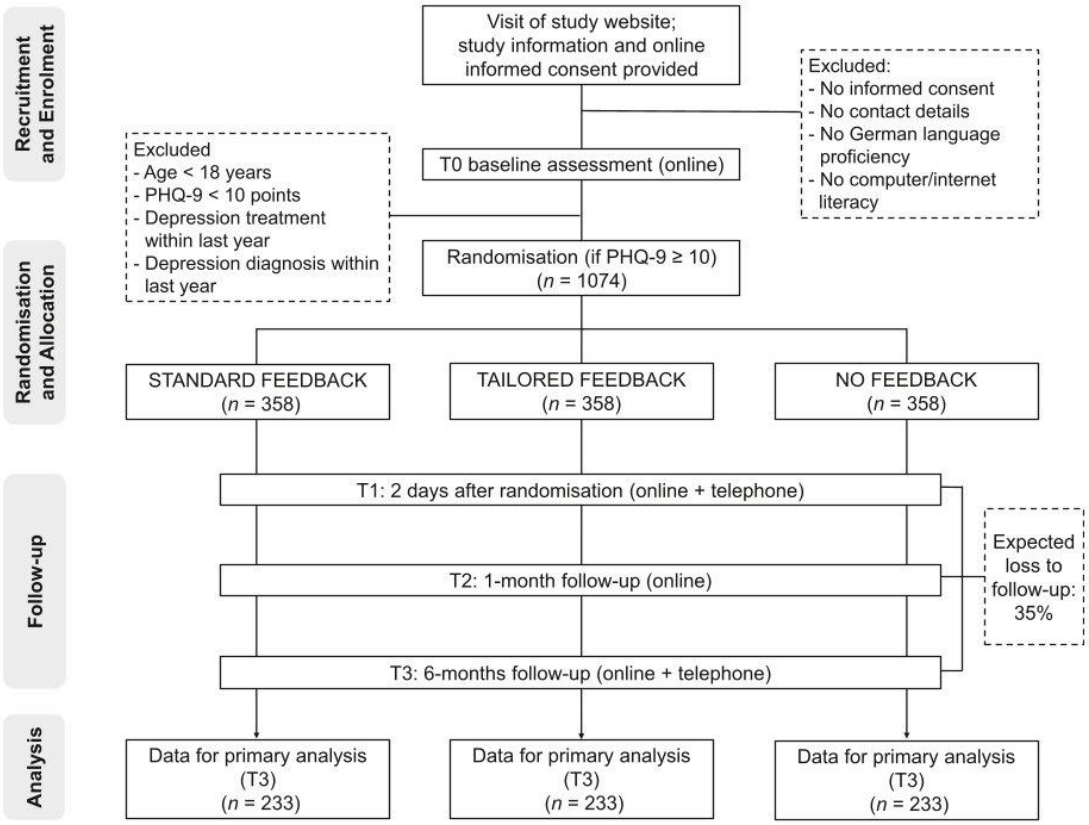


Fig. 1. Flow chart of the DISCOVER trial according to the SPIRIT 2010 statement.
Note. PHQ-9 = Patient-Health-Questionnaire-9.

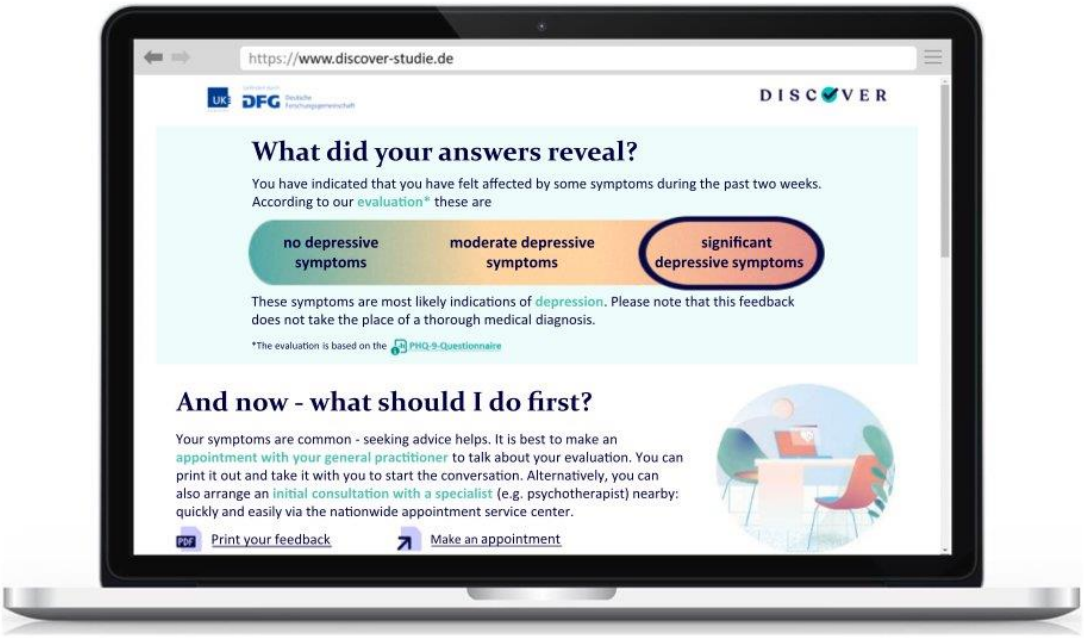


Fig. 2. Standard feedback: First screen as displayed on the DISCOVER study website (English translation).

2.5. Sample size

Based on the results of the preceding DEPSCREEN-INFO trial (Löwe et al., 2016), the study is powered to detect a small mean difference (Cohen's $f = 0.118$) in the primary outcome (depression severity) in any pairwise comparison between all three study arms. The calculation is based on a global one-way ANCOVA adjusted for baseline depression severity, with an alpha of 0.05 (two-sided) and a power of 80%. It results in a needed sample size of $n = 233$ participants per group (PASS, 2008). To allow for an estimated drop out of 35% (c.f. Christensen et al., 2009), 358 participants per group will be recruited (1074 in total).

2.6. Study arms

After completing the PHQ-9 depression screening questionnaire at T0, all eligible participants who score 10 points or higher will be directly randomised into one of the three study arms. Independent of the study arm, all participants will be provided with a 'thank you'-note and information on further follow-up procedures.

2.6.1. No feedback

This study arm serves as a passive control condition. The participants will not get any feedback on their screening result.

2.6.2. Standard feedback

Participants in this study arm will receive standardised feedback comprising the following four sections: (a) the *depression screening result*, (b) a *note to seek diagnostic consultation* by a health professional, (c) brief general *information on depression*, and (d) *information on depression treatment* (based on the German National Clinical Practice Guideline for Unipolar Depression; DGPPN et al., 2015). In line with the Common-Sense Model of Self-Regulation (Leventhal et al., 2003, 2016), the feedback content is designed to trigger adaptive illness beliefs such as an adequate illness identity, a coherent understanding of the condition, and optimistic control expectations. These, in turn, should guide patient engagement in functional health behaviour such as help-seeking and depression care.

The feedback intervention was developed in a multistage process. First, the underlying feedback version used in the preceding DEPSCREEN-INFO trial was subjected to re-evaluation and updating in several focus groups, involving patient representatives with depressive disorder (Seeralan et al., 2020). Based on the results of this qualitative study, needs and preferences of the target group could be assessed and implemented, resulting in the feedback version used in the currently running GET.FEEDBACK.GP trial (Kohlmann et al., 2020; see Supplemental Fig. I). For the use in DISCOVER, a digital art/graphic agency (Wood Agency, Hamburg) further adapted the feedback material to the possibilities of internet-based presentation. Namely, the present version is extended by (animated) graphic elements, adaptively available further information on specific contents, direct links to referenced health or social services (e.g. online therapies, self-help groups), and the possibility to download the feedback form as a pdf-file that includes the active links from the website. Throughout the process, the selection of content, design, and language was aligned to the current evidence on patients' needs in technology-based mental health interventions (e.g. Bakker et al., 2016; Hadjistavropoulos et al., 2018; Rozbroj et al., 2014; Torous et al., 2018).

Fig. 2 depicts an excerpt of the feedback screen as displayed in the desktop version (see Supplemental Fig. II, for the complete version). For smaller devices such as tablets and smartphones, the content is displayed in responsive design (i.e. the design automatically adapts to the size and type of the output device).

2.6.3. Tailored feedback

In order to trigger more salience, the content of the STANDARD FEEDBACK version is tailored to participants' characteristics as follows: First, the presentation of the screening result is framed according to participants' individual *symptom profiles* (e.g., 'You have indicated that you had *low spirits*, *sleep disturbances*, and *loss of energy* during the past two weeks.', see Fig. 3). Second, the note to seek further diagnostic consultation is matched to participants' *specialist preferences* (general practitioner vs. mental health professional). Third, the information on depression is tailored to participants' *symptom profiles* (e.g. 'Typical symptoms of depression are for example *low spirits* and *sleep disturbances*.')

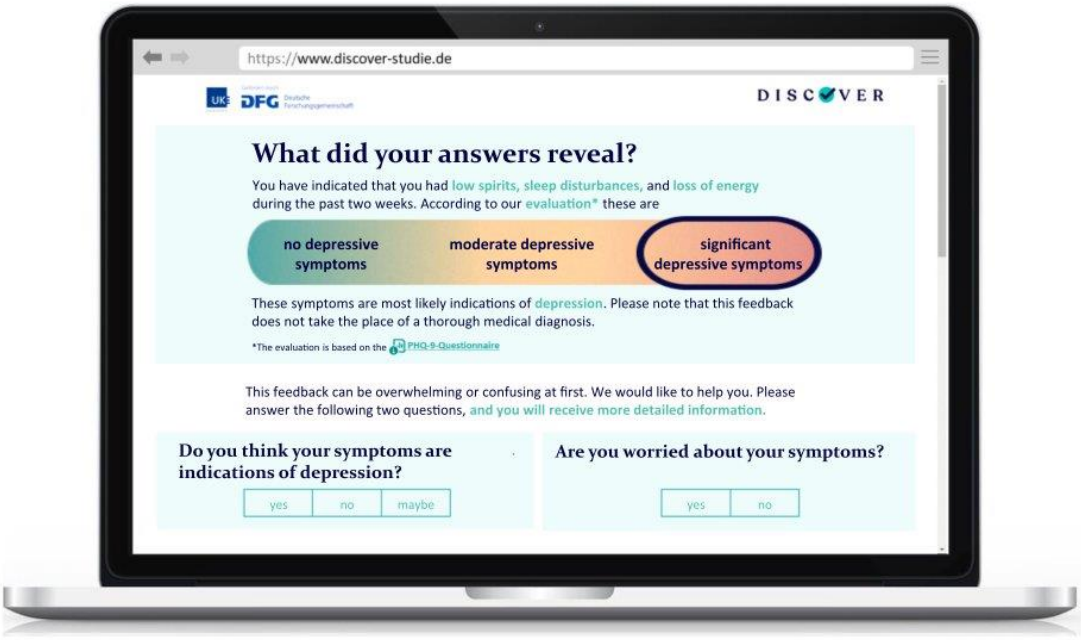


Fig. 3. Tailored feedback: First screen as displayed on the DISCOVER study website (English translation).

and their symptom *causal attributions* (e.g. 'Triggers are for example *stress with the partner*, *negative thinking patterns*, or a *physical illness*.'). Lastly, treatment options and help seeking advices are adapted to participants' *health insurance providers* and *local residency* in Germany (e.g. by providing links to self-help groups located nearby or to online therapies which are covered by the participant's health insurance provider).

Additionally, directly after being provided with the screening result, participants are asked the following two questions: 'Do you think your symptoms are indications of depression?' and 'Do you worry about your symptoms?' (see Fig. 3). According to participants' answers, the following three feedback sections are arranged in a differing order. If participants indicate assigning their symptoms to depression and worrying about them, the information on depression treatment is prefixed to the general information on depression, resulting in the following order: (b) note to seek diagnostic consultation, (c) information on depression treatment, (d) information on depression. If participants do not think that their symptoms relate to depression and/or do not worry about them, the information on depression is prefixed to the other sections, leading to the following order: (b) information on depression, (c) note to seek diagnostic consultation, (d) treatment information. Further, dependent on the combination of answers, the information on depression and the note to seek diagnostic consultation are phrased differently and are extended by information on depression prevalence and negative consequences of depression, both tailored to participants' *risk profile* (e.g. 'Depression is common, and particularly *people with diabetes* are often affected.', and 'In the long term, depressive symptoms have negative consequences – for example they can worsen the course of *diabetes*.'). Examples for the resulting feedback versions for all combinations of answers can be found in Supplemental Fig. III.

2.7. Outcomes

The primary outcome of the study will be self-reported depression severity (Patient Health Questionnaire-9) 6 months after screening. Secondary outcomes are guideline-based depression care (i.e. proportion of individuals treated according to the German depression guideline), depression-related help-seeking behaviour (i.e. proportion of individuals seeking formal/informal help), health-related quality of life, anxiety severity, somatic symptom severity, and adverse events, all at 6 months after screening, as well as depression severity and intervention acceptance, both at 1 month after screening. Further, 6 months after screening a health economic evaluation will be conducted based on direct costs (healthcare utilisation), indirect costs (productivity loss), and health-related quality of life. Corresponding measures are described in Section 2.8.

2.8. Data collection and measures

Data collection will be scheduled at baseline (before randomisation: T0; 2 days after randomisation: T1), and at 1-month (T2) and 6-months follow-up (T3). Assessments will comprise online self-report questionnaires (T0–T3) as well as clinical telephone interviews (T1 and T3 only). The baseline assessment is split into T0 and T1 two days later for two reasons: (a) to reduce potential recall effects from the PHQ-9 assessment at T0 to subsequent clinical interviews, and (b) to minimise participant burden and promote survey completion at T0. The latter is justified by the fact that only retrospective measures that are unlikely to be immediately influenced by the intervention (e.g. healthcare utilisation in the past 6 months) are assessed subsequently. To promote retention, email invitations to the online surveys will include information highlighting the importance of follow-up assessments and email reminders will be sent to participants at regular intervals if their surveys stay incomplete (up to 5, 7, and 10 reminders at T1, T2 and T3, respectively). All procedures will be managed computerised.

All measures will be entered into electronic data capture systems. The system for self-report data is implemented in the study website and shows one questionnaire (desktop version) or one question (smartphone version) per screen. It checks for completeness of questionnaires before submitting, allows participants to change their answers, and uses adaptive questioning to reduce the complexity of questionnaires, if applicable. In order to potentially identify invalid entries, all online surveys will comprise the following two questions as validity checks: (a) 'Have you answered the questions for yourself?' and (b) 'Have you answered the questions seriously?'

Table 1
Measures and assessment time points.

Measures	T0	T1	T2	T3
Primary outcome				
Depression severity, PHQ-9	x		x	x[a]
Secondary outcomes/process measures				
Guideline-based depression care (e.g. depression diagnosis, psychotherapy, medication)				x
Depression-related help-seeking behaviour (e.g. seeking information about depression)				x
Anxiety severity, GAD-7	x			x
Somatic symptom severity, SSS-8	x			x
Health-related quality of life, EQ-5D-5L	x			x
Healthcare utilisation and productivity loss, CSSRI		x		x
Intervention acceptance, USE			x	x
Illness beliefs, Brief IPQ	x		x	x
Intervention adherence	x			x[b]
Critical life events				x[b]
Depression diagnosis, SCID		x[b]		x[b]
Adverse events				x[b]
Website use	x	x	x	x
Characteristics				
Sociodemographic data	x			
Medical data	x			
Risk factors for depression onset	x			

Note. T0 = before randomisation; T1 = 2 days after randomisation; T2 = 1-month follow-up, T3 = 6-months follow-up; PHQ-9 = Patient Health Questionnaire-9; CSSRI = Client Sociodemographic and Service Receipt Inventory; EQ-5D-5L = EuroQol-5D 5-L; GAD-7 = Generalized Anxiety Disorder-7; SSS-8 = Somatic Symptom Scale-8; SCID = Structured Clinical Interview for DSM-5 Disorders; USE = Usefulness Scale for Patient Information Material; Brief IPQ = Brief Illness Perception Questionnaire.
[a] Primary outcome.
[b] Measures assessed via telephone interview.

Table 1 shows an overview of all measures and corresponding assessment time points.

2.8.1. Depression severity

Depression severity will be assessed by the German version of the Patient Health Questionnaire-9 (PHQ-9; Kroenke et al., 2001; Löwe et al., 2004a,b). The PHQ-9 consists of 9 items covering all major depression symptom criteria as stated in the DSM-5. Each item refers to the past two weeks and is scored on a 4-point Likert scale (0–3), resulting in a total score ranging from 0 to 27. The PHQ-9 is among the most frequently used and best validated self-report depression questionnaires: it has good psychometric properties, is sensitive to change and responsive to treatment (Kroenke et al., 2001; Löwe et al., 2004b). When delivered online, it has shown to have a good inter-format reliability to the paper version (Erbe et al., 2016). With regard to depression screening, the PHQ-9 (cut-off of 10 points) is recommended as the most suitable instrument compared with others in a recent meta-analysis (Miller et al., 2021), showing high sensitivity (0.88) and specificity (0.85; Levis et al., 2019). Further, the PHQ-9 is recommended for depression screening also by national clinical expert associations such as the US Preventive Services Task Force (Siu et al., 2016) and the German National Clinical Practice Guideline for Unipolar Depression (DGPPN et al., 2015).

2.8.2. Guideline-based depression care and depression-related help-seeking behaviour

In absence of a standardised measure for evaluating depression-related health behaviour and depression care according to the German national guideline, these will be assessed via a self-developed questionnaire. The questionnaire comprises guideline-based depression care (e.g. depression diagnosis by a health professional, psychotherapy, medication), formal help-seeking (e.g. contacting any health professional), and informal help-seeking (e.g. seeking information, doing exercise), as well as the perceived helpfulness, respectively. Items are developed based on recommendations of the German National Clinical Practice Guideline for Unipolar Depression (DGPPN et al., 2015) and extended by questions in an open format. For formal help-seeking and depression care, the time point (in months after the intervention) and specific characteristics (e.g. type of professional contacted) will be assessed. In a similar version, these questions have been successfully tested in the preceding DEPSCREEN-INFO trial (Löwe et al., 2016).

2.8.3. Anxiety severity

Anxiety severity during the past two weeks will be assessed with the 7-item Generalized Anxiety Disorder Scale (GAD-7; Spitzer et al., 2006), which is widely used for this purpose and well validated in its German version (Löwe et al., 2008).

2.8.4. Somatic symptom severity

The Somatic Symptom Scale-8 (SSS-8; Gierk et al., 2014) will be used to assess somatic symptom severity. The questionnaire consists of 8 items that reflect common somatic symptoms in primary care and refer to the past two weeks. It has good psychometric properties and is sensitive to change (Gierk et al., 2017).

2.8.5. Health-related quality of life

The widely used 5-level version of the EuroQol-5D (EQ-5D-5L; Herdman et al., 2011) will be used to assess health-related quality of life. The generic questionnaire comprises 5 items relating to the following dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Furthermore, a visual analogue scale records overall perceived health status. The instrument is widely used and responsive to treatment (Sobocki et al., 2007). Preference-based utilities derived from the EQ-5D-5L (Ludwig et al., 2018) will be used to calculate quality-adjusted life-years (QALYs) for the health economic evaluation. This approach is evaluated suitable for this purpose in the field of depression (Lamers et al., 2006; Sapin et al., 2004).

2.8.6. Healthcare utilisation and productivity loss

Healthcare utilisation and productivity loss will be assessed with an adapted version of the Client Sociodemographic and Service Receipt Inventory (CSSRI; Chisholm et al., 2000). It registers the use of healthcare services (e.g. hospital stays, health professional contacts), medication (e.g. type of drug, dosage level), and work loss days (e.g. hospital days, absenteeism) during the past 6 months.

2.8.7. Depression diagnosis

To validate the suspected diagnosis of depression indicated by the PHQ-9 depression screening, the depression related modules of the Structured Clinical Interview for DSM-5 Disorders (SCID-5-CV; Beesdo-Baum et al., 2019) will be conducted. The SCID enables a reliable, valid and efficient assessment of depressive disorders according to DSM-5 criteria. Interviews will be conducted via telephone, which has demonstrated high inter-rater reliability when compared to face-to-face interviews (Crippa et al., 2008). To ensure validity and reliability, the assessors (BSc or MSc Psychology) will undergo a standardised training and will be supervised by an experienced psychotherapist (PhD).

2.8.8. Intervention acceptance

The Usefulness Scale for Patient Information Material (Holzel et al., 2015) will be used to assess the acceptance of the feedback intervention. The original instrument consists of 9 items assessing cognitive, emotional and behavioural aspects of usefulness and has excellent psychometric properties. For the present study, one item was added to assess whether the feedback information appeared trustworthy. To assess the acceptance of depression screening, directly after filling in the PHQ-9 the following dichotomous items will be added: ‘Answering these questions… (a) bothered/did not bother me, (b) was easy/complicated, (c) was too/was not too time-consuming’, (d) ‘Answering these questions on the internet is a problem/no problem’, (e) ‘Answering these

questions at a general practitioner would be a problem/no problem', and (f) 'In a similar life situation, would you answer these questions again on the internet?'. Moreover, in the telephone interviews (T3) participants will be asked two open questions regarding the perceived helpfulness of the feedback ('Did you find the feedback helpful (why/why not)?') and the perceived helpfulness of internet-based depression screening ('Do you think an internet-based questionnaire such as the one used in the DISCOVER study is helpful to improve early detection of psychological distress (why/why not?)?').

2.8.9. Illness beliefs

Illness beliefs regarding depressive symptoms will be measured with a modified version of the well validated Brief Illness Perception questionnaire (Brief IPQ, Broadbent et al., 2006). The Brief IPQ is based on the Common-Sense Model of Self-Regulation (Leventhal et al., 2016) and covers causal, cognitive and emotional representations of an illness (identity, coherence, causes, consequences, timeline, personal and treatment control, and worry). As the target population of non-diagnosed individuals might not associate their symptoms with an 'illness', this term will be replaced with 'symptoms' throughout the questionnaire. The item assessing illness identity will be replaced by the dichotomous questions 'Can you imagine suffering from depression?' (T1 and T3) and 'Can you imagine having suffered from depression within the last six months?' (T3 only) as well as the open question 'In your own words – how would you describe your mental health in the last six months? Do you think you suffered from depression?' (T3 telephone interview). Further, the open question for the causal representations will be complemented by a listing of potential causes of depressive symptoms adopted from the Beliefs about Depression Questionnaire (Lynch et al., 2011).

2.8.10. Adverse events

To estimate possible unintended adverse events of the feedback intervention, at T3 participants will be asked about the occurrence of any negative event that is attributed to the trial with an open question.

2.8.11. Critical life events

Three open questions assessing relevant positive and negative critical life events will be asked at T3: 'Within the last six months, … (a) Did you experience life events that positively influenced your mood?, (b) Did you experience life events that negatively influenced your mood?, (c) and What has been particularly helpful to you in times when you have been feeling bad?'.

2.8.12. Intervention adherence

Intervention adherence will be assessed by the item 'Please indicate to what extent you have read the feedback with the corresponding information.' and the following response options: '100%', '90%', '75%', '50%', '25%', '10%', and '0%'. To complement this self-report data, the system will also track technical data on feedback use (e.g. time spent on the screen, documents downloaded).

2.8.13. Website use

In order to obtain additional measures for acceptability and usability of the applications as well as to monitor and potentially improve processes during the trial (e.g. recruitment success, problems with usability), technical data on website (including questionnaire) use will be recorded by the system (e.g. hits per page, usage time).

2.8.14. Characteristics

Participant characteristics recorded at T0 will include sociodemographic data (e.g. age, gender, education, family status, rural/urban area living, local residency, health insurance provider), risk factors for depression onset (e.g. chronic somatic comorbidities, pregnancy, alcohol and nicotine consumption), and medical data (diagnosis of and treatments for depression).

2.9. Data storage and management

To ensure participants' data safety, study data and personal data will be stored in separate data bases. Security of data transmission from data capture software to data bases is guaranteed by a TLS-encrypted connection. For the duration of the study, a University-hosted pseudomisation service (Mainzelliste; Lablans et al., 2015) will enable the temporary connection of personal with study data, which is necessary for the follow-up assessments. Compliance of these procedures with the security requirements enforced by the European General Data Protection Regulation as well as German law is ensured. Constant monitoring and backups of data as well as password-restricted access will be ensured by an external IT company (Timo Stolz, Berlin).

In accordance with the German Research Foundation guidelines for the handling of research data, the de-identified data will be saved for at least 10 years (i.e. analysable data set, protocol, statistical analysis plan and statistical programming code). Data sharing will follow the FAIR Data Principles (Findable, Accessible, Interoperable and Reusable) to maximise transparency and scientific reproducibility. The data management plan will (a) ensure long-term accessibility, (b) deliver a comprehensive, reliable view of data and (c) provide a future-proof solution for international healthcare interoperability.

2.10. Data analysis

Data analysis will be conducted by an independent statistician from the Department of Medical Biometry and Epidemiology who will be blind to the research hypotheses. All pre-specified analyses will be conducted according to the intention-to-treat (ITT) principle, i.e. including all participants randomised. In addition to the following description, planned analyses will be specified in accordance with the current statistical recommendations of the European Medicine Agency in a statistical analysis plan that will be signed by the principal investigator and the responsible statistician before breaking the blinding.

A multilevel model incorporating the participants as random terms will be applied to the repeated measures in the same participant, including the factor group and the baseline value for adjustment. The primary analysis will be performed within the framework of this model as an ANCOVA of the PHQ-9 change scores (T0 to T3-difference), with subsequent pairwise comparisons of interventions by test of the corresponding contrasts. Each test will be performed at a two-sided level of alpha = 0.05. This closed testing principle will ensure a family-wise error level of 5%. The multilevel modelling approach limits the bias when handling missing data even in the case of not missing at random (NMAR). However, alternative missing data mechanisms will be applied as a sensitivity check to examine the stability of the results. No subgroup analyses are pre-specified.

For the health-economic evaluation, the cost-effectiveness of the feedback interventions compared to no feedback will be determined. For this, incremental cost-effectiveness ratios (ICER) will be calculated as the difference in mean costs divided by the difference in mean QALYs between each of the two intervention groups and the control group. Net benefit regressions will be conducted to determine the uncertainty of the point estimates and to adjust for potential baseline differences and confounders (Briggs et al., 2002). To show the intervention's probability of being cost-effective at different willingness-to-pay margins in comparison to each of the two comparators, cost-effectiveness acceptability curves will be derived.

3. Discussion

The high prevalence of undetected major depression underscores the relevance of new approaches that ideally target both its early detection and resolution. With the DISCOVER RCT, we address this by testing the efficacy of automated feedback after internet-based depression screening.

The primary outcome of the trial will be depression severity six months after screening. Based on the results of our preceding trial (Löwe et al., 2016), we expect the feedback intervention to have a small effect on depression severity. Further, we expect the tailored feedback version to amplify the effect of the standard version to a small extent. Although being small in magnitude, this effect size is clinically relevant as it intents to address a so far un-diagnosed population that, until now, falls outside the scope of any form of depression care. Therefore, we believe that the small effect at the individual level leads to a substantial effect at the larger population level.

Whereas our preceding RCTs DEPSCREEN-INFO and GET.FEEDBACK.GP investigate(d) feedback after depression screening in patients with coronary heart disease (Löwe et al., 2016) and in primary care (Kohlmann et al., 2020), the internet-based format of DISCOVER allows for a wider reach and may also attract people who are reluctant to seek traditional health services, but use the internet for mental health information (c.f. Berger et al., 2005). Addressing this large population of affected individuals outside of the medical system, the results of DISCOVER will expand on those of our preceding trials.

Furthermore, the DISCOVER RCT will allow for a deeper understanding of the early detection and resolution processes. So far, it is unclear how exactly informing patients about their screening result translates into improved depression severity (Löwe et al., 2016). Also, there appears to be a lack of knowledge on how to get undetected individuals into treatment. The comprehensive examination of process variables such as illness beliefs and depression-related help-seeking behaviour could be a contribution in this regard. Depending on the ultimately reached recruitment rate and the resulting power, also process focussed analyses could be conducted. Results regarding the underlying processes of feedback after depression screening could improve the refinement and development of further feedback as well as other interventions targeting patient engagement in early depression detection.

With regard to practical implication, the brevity of the feedback intervention makes it suitable, when further validated, for widespread implementation in different contexts: potential modes of dissemination could target for example mental health-related websites (e.g. of health insurances, doctors' practices), but also social media (e.g. forums on mental health topics) or websites of community institutions with a high reach (e.g. universities). Taking into account these aspects, the internet-based feedback intervention could be a worthwhile contribution to improving early detection and resolution of depression.

3.1. Strengths and limitations

Testing the feedback intervention in an internet-based trial involves possible limitations, which we try to overcome using the following approaches. First, the trial relies on self-selection of participants, and internet-savvy individuals and/or those interested in mental health might be overrepresented. To minimise this potential bias, we will monitor sample characteristics during recruitment and will adapt strategies appropriately (e.g. by targeted advertisement and by involving a population wide survey panel). Second, drop-out in internet-based interventions can be moderately to high (Melville et al., 2010), which can lead to reduced power of analyses. We will approach this problem in different ways. To promote retention, the importance of follow-ups will be highlighted in all study instructions and participants will receive automated email reminders. Furthermore, to handle inevitable drop-out, we anticipated a drop-out rate of 35% in the sample size calculation and will further analyse data on an ITT basis using adequate mechanisms for handling missing data. Third, part of the intervention effect might be due to the feedback intervention increasing individuals' awareness of their symptoms. It cannot be ruled out that the questionnaires and interviews at baseline might trigger a similar process. Due to randomisation this effect should occur in all three study arms. However, as it could be confounded with the intervention effect, this might lead to the resulting efficacy being underestimated as compared to real-life conditions. Lastly, some researchers argue that depression screening by self-report questionnaires might pose the risk of over-diagnosis of depression, which again might lead to over-treatment (Thombs et al., 2014). To account for this, we investigate possible over-treatment due to our intervention by verifying suspected depression diagnosis with a gold standard clinical interview (SCID) and by recording participants' healthcare use six months after the intervention.

Several strengths of the DISCOVER trial should be highlighted as well. First, the feedback intervention is a result of an elaborated multi-stage development process, which combined strengths and perspectives of different domains: clinical, research, and IT/graphic design expertise, empirical evidence, and first-hand patients' needs and preferences (Seeralan et al., 2020). Second, the selection of a broad range of further outcomes (depression care and help-seeking behaviour, additional clinical outcomes, intervention acceptance, illness beliefs, adverse events, and the health-economic evaluation) allows for a comprehensive trial evaluation. Third, DISCOVER extensively exploits the potential of technology-based trial design - for example, by (a) automated randomisation and allocation to ensure standardisation of trial conduction, (b) interactively tailoring the feedback intervention to participant characteristics to increase its suitability, (c) automated management of assessments and reminders to improve retention, (d) impeding double or 'fake' entries by several security checks, and (e) well-designed online questionnaire administration (e.g. adaptive questioning) to minimise participant burden. Lastly, the assessment of technical data on participants' website and questionnaire use allows for a thorough investigation of user behaviour, which could contribute to the evidence on optimal clinical trial design of internet-based interventions in the future.

3.2. Conclusion

Taken together, the DISCOVER RCT is well designed to yield comprehensive information on how automated feedback after internet-based screening could improve early detection and resolution of depression. If the results show that automated feedback after internet-based depression screening can reduce depression severity, the intervention could be easily and widely disseminated. The trial could further contribute to an understanding of the help-seeking behaviour processes initiated after internet-based depression screening with automated feedback, which could inform further research and practical implementation. Therefore, the results of the DISCOVER RCT will show whether, and if so, how automated feedback after internet-based screening can improve the early detection and resolution of undetected depression.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.invent.2021.100435.

CRediT authorship contribution statement

FS wrote the first draft of this manuscript under supervision of SK. All authors revised the draft critically for important intellectual content and contributed substantially to the conception of the study. The applicants of the DISCOVER trial are SK (principal investigator), BL, KW & H-HK. KW and AZ are the trial statisticians and contributed to the analysis aspect of the protocol. All authors gave approval of the version published.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank Marieke Volkmann and Nele Schade for extensively pilot-testing the website and the data entry systems, as

well Marieke Volkmann for providing help with creating the figures for the manuscript.

Funding

This work was funded by the German Research Foundation (grant number: 424162019).

Compliance with ethical standards

The trial was approved by the Ethics Committee of the Medical Chamber Hamburg in July 2019 (reference number: PV7039).

References

Andersson, G., 2016. Internet-delivered psychological treatments. Annu. Rev. Clin. Psychol. 12, 157–179. https://doi.org/10.1146/annurev-clinpsy-021815-093006.

Andersson, G., Titov, N., 2014. Advantages and limitations of internet-based interventions for common mental disorders. World Psychiatry 13 (1), 4–11. https://doi.org/10.1002/wps.20083.

Bakker, D., Kazantzis, N., Rickwood, D., Rickard, N., 2016. Mental health smartphone apps: review and evidence-based recommendations for future developments. JMIR Ment. Health 3 (1). https://doi.org/10.2196/mental.4984.

Beesdo-Baum, K., Zaudig, M., Wittchen, H.U., Williams, Janet B.W., Karg, Rhonda S., Spitzer, Robert L., 2019. In: Von Michael, B. First (Ed.), SCID-5-CV Strukturiertes Klinisches Interview für DSM-5-Störungen–Klinische Version: Deutsche Bearbeitung des Structured Clinical Interview for DSM-5 Disorders–Clinician Version. Hogrefe.

Berger, M., Wagner, T.H., Baker, L.C., 2005. Internet use and stigmatized illness. Soc. Sci. Med. 61 (8), 1821–1827. https://doi.org/10.1016/j.socscimed.2005.03.025.

Boerema, A.M., Kleiboer, A., Beekman, A.T., van Zoonen, K., Dijkshoorn, H., Cuijpers, P., 2016. Determinants of help-seeking behavior in depression: a cross-sectional study. BMC Psychiatry 16 (1), 78. https://doi.org/10.1186/s12888-016-0790-0.

Boutron, I., Altman, D.G., Moher, D., Schulz, K.F., Ravaud, P., 2017. CONSORT statement for randomized trials of nonpharmacologic treatments: a 2017 update and a CONSORT extension for nonpharmacologic trial abstracts. Ann. Intern. Med. 167 (1), 40–47. https://doi.org/10.7326/m17-0046.

Briggs, Andrew, O'Brien, Bernie, Blackhouse, Gordon, 2002. Thinking outside the box: recent advances in the analysis and presentation of uncertainty in cost-effectiveness studies. Annual Review of Public Health 23, 377–401. https://doi.org/10.1146/annurev.publhealth.23.100901.140534.

Broadbent, E., Petrie, K.J., Main, J., Weinman, J., 2006. The brief illness perception questionnaire. J. Psychosom. Res. 60 (6), 631–637. https://doi.org/10.1016/j.jpsychores.2005.10.020.

Busch, M., Maske, U., Ryl, L., Schlack, R., Hapke, U., 2013. Prevalence of depressive symptoms and diagnosed depression among adults in Germany - results of the german health interview and examination survey for adults (DEGS1). Epidemiol. Gesundheitsberichterstattung 56 (5), 733–739. https://doi.org/10.1007/s00103-013-1688-3.

Chan, A.W., Tetzlaff, J.M., Altman, D.G., Laupacis, A., Gotzsche, P.C., Krleza-Jeric, K., Moher, D., 2013. SPIRIT 2013 statement: defining standard protocol items for clinical trials. Ann. Intern. Med. 158 (3), 200–207. https://doi.org/10.7326/0003-4819-158-3-201302050-00583.

Chisholm, D., Knapp, M.R.J., Knudsen, H.C., Amaddeo, F., Gaite, L., Van Wijngaarden, B., 2000. Client socio-demographic and service receipt inventory-european version: development of an instrument for international research. Br. J. Psychiatry Suppl. 177 (39), s28–s33. https://doi.org/10.1192/bjp.177.39.s28.

Chisholm, D., Sweeny, K., Sheehan, P., Rasmussen, B., Smit, F., Cuijpers, P., Saxena, S., 2016. Scaling-up treatment of depression and anxiety: a global return on investment analysis. Lancet Psychiatry 3 (5), 415–424. https://doi.org/10.1016/S2215-0366(16)30024-4.

Christensen, H., Griffiths, K.M., Farrer, L., 2009. Adherence in internet interventions for anxiety and depression: systematic review. J. Med. Internet Res. 11 (2), e13 https://doi.org/10.2196/jmir.1194.

Crippa, J.A., de Lima Osório, F., Del-Ben, C.M., Filho, A.S., da Silva Freitas, M.C., Loureiro, S.R., 2008. Comparability between telephone and face-to-face structured clinical interview for DSM-IV in assessing social anxiety disorder. Perspect. Psychiatr. Care 44 (4), 241–247. https://doi.org/10.1111/j.1744-6163.2008.00183.x.

DGPPN, BÄK, KBV, AWMF, 2015. S3-Leitlinie/Nationale VersorgungsLeitlinie Unipolare Depression – Langfassung (edition 2, version 5). Retrieved from: https://www.awmf.org/leitlinien/detail/ll/nvl-005.html.

Ebert, D.D., Cuijpers, P., Muñoz, R.F., Baumeister, H., 2017. Prevention of mental health disorders using internet-and Mobile-based interventions: a narrative review and recommendations for future research. Front. Psychiatry 8. https://doi.org/10.3389/fpsyt.2017.00116.

Eichenberg, Christiane, Wolters, Carolin, Brähler, Elmar, 2013. The internet as a mental health advisor in Germany—results of a national survey. PloS one 8 (11), e79206. https://doi.org/10.1371/journal.pone.0079206.

Erbe, D., Eichert, H.C., Rietz, C., Ebert, D., 2016. Interformat reliability of the patient health questionnaire: validation of the computerized version of the PHQ-9. Internet Interv. 5, 1–4. https://doi.org/10.1016/j.invent.2016.06.006.

Eysenbach, G., Group, C.-E, 2011. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J. Med. Internet Res. 13 (4), e126-e126 https://doi.org/10.2196/jmir.1923.

Fichter, M.M., Quadflieg, N., Fischer, U.C., Kohlboeck, G., 2010. Twenty-five-year course and outcome in anxiety and depression in the upper bavarian longitudinal community study. Acta Psychiatr. Scand. 122 (1), 75–85. https://doi.org/10.1111/j.1600-0447.2009.01512.x.

Ghio, L., Gotelli, S., Marcenaro, M., Amore, M., Natta, W., 2014. Duration of untreated illness and outcomes in unipolar depression: a systematic review and meta-analysis. J. Affect. Disord. 152, 45–51. https://doi.org/10.1016/j.jad.2013.10.002.

Gierk, B., Kohlmann, S., Hagemann-Goebel, M., Löwe, B., Nestoriuc, Y., 2017. Monitoring somatic symptoms in patients with mental disorders: sensitivity to change and minimal clinically important difference of the somatic symptom scale – 8 (SSS-8). Gen. Hosp. Psychiatry 48, 51–55. https://doi.org/10.1016/j.genhosppsych.2017.07.002.

Gierk, B., Kohlmann, S., Kroenke, K., Spangenberg, L., Zenger, M., Brähler, E., Löwe, B., 2014. The somatic symptom Scale–8 (SSS-8): a brief measure of somatic symptom burden. JAMA Intern. Med. 174 (3), 399–407. https://doi.org/10.1001/jamainternmed.2013.12179.

Gilbody, S., Sheldon, T., House, A., 2008. Screening and case-finding instruments for depression: a meta-analysis. CMAJ 178 (8), 997–1003. https://doi.org/10.1503/cmaj.070281.

Hadjistavropoulos, H.D., Faller, Y.N., Klatt, A., Nugent, M.N., Dear, B.F., Titov, N., 2018. Patient perspectives on strengths and challenges of therapist-assisted internet-delivered cognitive behaviour therapy: using the patient voice to improve care. Community Ment. Health J. 54 (7), 944–950. https://doi.org/10.1007/s10597-018-0286-0.

Herdman, M., Gudex, C., Lloyd, A., Janssen, M., Kind, P., Parkin, D., Badia, X., 2011. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual. Life Res. 20 (10), 1727–1736. https://doi.org/10.1007/s11136-011-9903-x.

Holzel, L.P., Ries, Z., Dirmaier, J., Zill, J.M., Kriston, L., Klesse, C., Bermejo, I., 2015. Usefulness scale for patient information material (USE) - development and psychometric properties. BMC Med. Inform. Decis. Mak. 15 https://doi.org/10.1186/s12911-015-0153-7.

Karyotaki, E., Riper, H., Twisk, J., Hoogendoorn, A., Kleiboer, A., Mira, A., Littlewood, E., 2017. Efficacy of self-guided internet-based cognitive behavioral therapy in the treatment of depressive symptoms: a meta-analysis of individual participant data. JAMA Psychiatry 74 (4), 351–359. https://doi.org/10.1001/jamapsychiatry.2017.0044.

Kohlmann, S., Lehmann, M., Eisele, M., Braunschneider, L.E., Marx, G., Zapf, A., Lowe, B., 2020. Depression screening using patient-targeted feedback in general practices: study protocol of the german multicentre GET.FEEDBACK.GP randomised controlled trial. BMJ Open 10 (9), e035973. https://doi.org/10.1136/bmjopen-2019-035973.

Kroenke, K., Spitzer, R.L., Williams, J.B.W., 2001. The PHQ-9. J. Gen. Intern. Med. 16 (9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x.

Lablans, M., Borg, A., Uckert, F., 2015. A RESTful interface to pseudonymization services in modern web applications. BMC Med. Inform. Decis. Mak. 15, 2. https://doi.org/10.1186/s12911-014-0123-5.

Lamers, L.M., Bouwmans, C.A., van Straten, A., Donker, M.C., Hakkaart, L., 2006. Comparison of EQ-5D and SF-6D utilities in mental health patients. Health Econ. 15 (11), 1229–1236. https://doi.org/10.1002/hec.1125.

Leventhal, H., Brissette, I., Leventhal, E., 2003. The common-sense model of self-regulation of health and illness. In: Cameron, L.D., Leventhal, H. (Eds.), The Self-Regulation of Health and Illness Behavior. Routledge, London, pp. 42–65.

Leventhal, H., Phillips, L.A., Burns, E., 2016. The common-sense model of self-regulation (CSM): a dynamic framework for understanding illness self-management. J. Behav. Med. 39 (6), 935–946. https://doi.org/10.1007/s10865-016-9782-2.

Levesque, D.A., Van Marter, D.F., Schneider, R.J., Bauer, M.R., Goldberg, D.N., Prochaska, J.O., Prochaska, J.M., 2011. Randomized trial of a computer-tailored intervention for patients with depression. Am. J. Health Promot. 26 (2), 77–89. https://doi.org/10.4278/ajhp.090123-QUAN-27.

Levis, B., Benedetti, A., Thombs, B.D., 2019. Accuracy of patient health Questionnaire-9 (PHQ-9) for screening to detect major depression: individual participant data meta-analysis. BMJ 365, l1476. https://doi.org/10.1136/bmj.l1476.

Löwe, B., Blankenberg, S., Wegscheider, K., König, H.-H., Walter, D., Murray, A.M., Kohlmann, S., 2016. Depression screening with patient-targeted feedback in cardiology: DEPSCREEN-INFO randomised clinical trial. Br. J. Psychiatry 1–8. https://doi.org/10.1192/bjp.bp.116.184168.

Löwe, B., Decker, O., Müller, S., Brähler, E., Schellberg, D., Herzog, W., Herzberg, P.Y., 2008. Validation and standardization of the generalized anxiety disorder screener (GAD-7) in the general population. Med. Care 46 (3), 266–274. https://doi.org/10.1097/MLR.0b013e318160d093.

Löwe, B., Spitzer, R.L., Gräfe, K., Kroenke, K., Quenter, A., Zipfel, S., Buchholz, C., Witte, S., Herzog, W., 2004. Comparative validity of three screening questionnaires for DSM-IV depressive disorders and physicians' diagnoses. J. Affect. Disord. 78 (2), 131–140. https://doi.org/10.1016/s0165-0327(02)00237-9.

Löwe, B., Unützer, J., Callahan, C.M., Perkins, A.J., Kroenke, K., 2004. Monitoring depression treatment outcomes with the patient health questionnaire-9. Med. Care 42 (12), 1194–1201. https://doi.org/10.1097/00005650-200412000-00006.

Ludwig, K., Graf von der Schulenburg, J.M., Greiner, W., 2018. German value set for the EQ-5D-5L. PharmacoEconomics 36 (6), 663–674. https://doi.org/10.1007/s40273-018-0615-8.

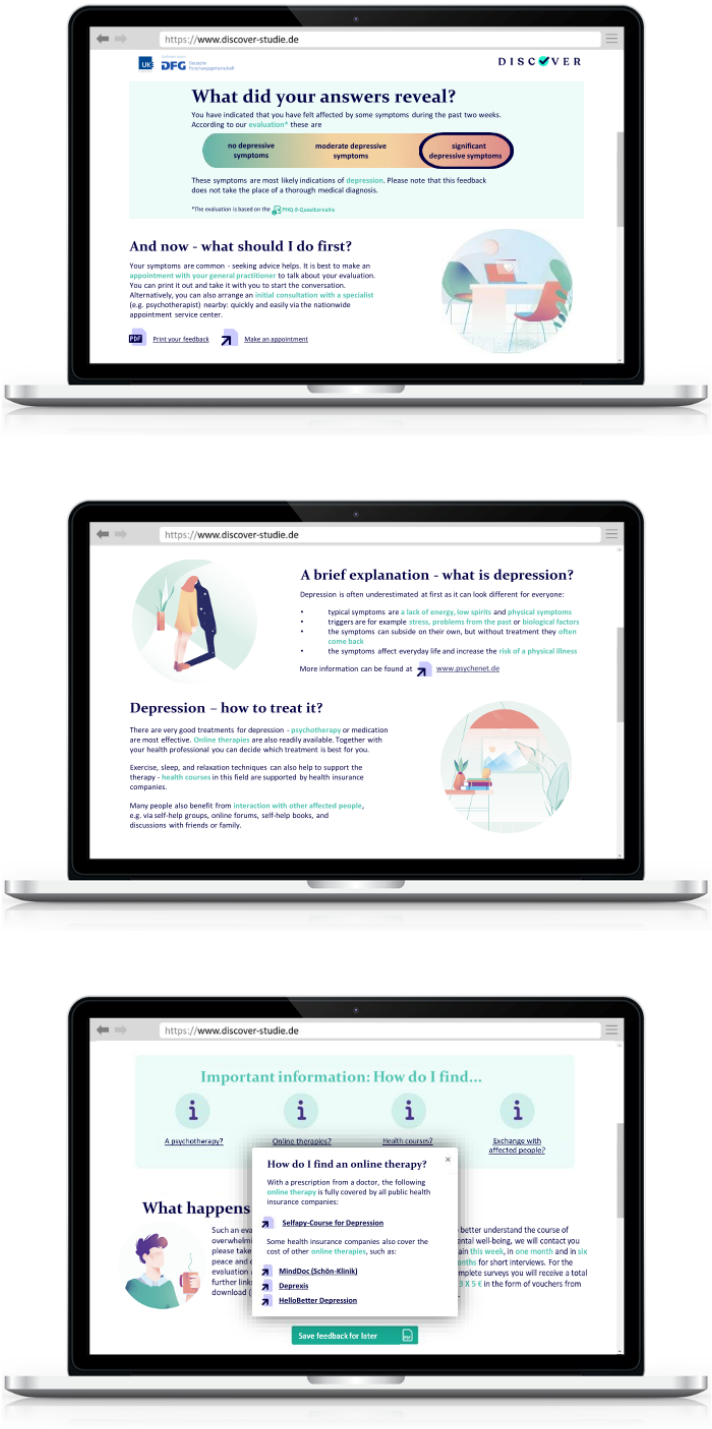
Lynch, J., Moore, M., Moss-Morris, R., Kendrick, T., 2011. Are patient beliefs important in determining adherence to treatment and outcome for depression? Development of

the beliefs about depression questionnaire. J. Affect. Disord. 133 (1–2), 29–41. https://doi.org/10.1016/j.jad.2011.03.019.
Melville, K.M., Casey, L.M., Kavanagh, D.J., 2010. Dropout from internet-based treatment for psychological disorders. Br. J. Clin. Psychol. 49 (Pt 4), 455–471. https://doi.org/10.1348/014466509x472138.
Miller, P., Newby, D., Walkom, E., Schneider, J., Li, S.C., Evans, T.-J., 2021. The performance and accuracy of depression screening tools capable of self-administration in primary care: a systematic review and meta-analysis. Eur. J. Psychiatry 35 (1), 1–18. https://doi.org/10.1016/j.ejpsy.2020.10.002.
Mitchell, A.J., Vaze, A., Rao, S., 2009. Clinical diagnosis of depression in primary care: a meta-analysis. Lancet 374 (9690), 609–619. https://doi.org/10.1016/S0140-6736(09)60879-5.
Moher, D., Hopewell, S., Schulz, K.F., Montori, V., Gøtzsche, P.C., Devereaux, P.J., Altman, D.G., 2010. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ 340, c869. https://doi.org/10.1136/bmj.c869.
Montgomery, P., Grant, S., Mayo-Wilson, E., Macdonald, G., Michie, S., Hopewell, S., Group, o. b. o. t. C.-S, 2018. Reporting randomised trials of social and psychological interventions: the CONSORT-SPI 2018 extension. Trials 19 (1), 407. https://doi.org/10.1186/s13063-018-2733-1.
Richards, D., Richardson, T., 2012. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin. Psychol. Rev. 32 (4), 329–342. https://doi.org/10.1016/j.cpr.2012.02.004.
Rohde, F., Franke, M., Sehili, Z., Lablans, M., Rahm, E., 2021. Optimization of the mainzelliste software for fast privacy-preserving record linkage. J. Transl. Med. 19 (1), 1–12. https://doi.org/10.1186/s12967-020-02678-1.
Rozbroj, T., Lyons, A., Pitts, M., Mitchell, A., Christensen, H., 2014. Assessing the applicability of E-therapies for depression, anxiety, and other mood disorders among lesbians and gay men: analysis of 24 web- and Mobile phone-based self-help interventions. J. Med. Internet Res. 16 (7), 84–94. https://doi.org/10.2196/jmir.3529.
Ryan, G.L., Skinner, C.S., Farrell, D., Champion, V.L., 2001. Examining the boundaries of tailoring: the utility of tailoring versus targeting mammography interventions for two distinct populations. Health Educ. Res. 16 (5), 555–566. https://doi.org/10.1093/her/16.5.555.
Sapin, C., Fantino, B., Nowicki, M.L., Kind, P., 2004. Usefulness of EQ-5D in assessing health status in primary care patients with major depressive disorder. Health Qual. Life Outcomes 2, 20. https://doi.org/10.1186/1477-7525-2-20.
Schomerus, G., Angermeyer, M.C., 2008. Stigma and its impact on help-seeking for mental disorders: what do we know? Epidemiol. Psichiatr. Soc. 17 (1), 31–37. https://doi.org/10.1017/s1121189x00002669.
Schulz, K.F., Altman, D.G., Moher, D., 2010. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMC Med. 8 (1), 18. https://doi.org/10.1186/1741-7015-8-18.
Seeralan, T., Harter, M., Koschnitzke, C., Scholl, M., Kohlmann, S., Lehmann, M., Brutt, A.L., 2020. Patient involvement in developing a patient-targeted feedback intervention after depression screening in primary care within the randomized controlled trial GET.FEEDBACK.GP. Health Expect. https://doi.org/10.1111/hex.13039.
Shah, R., Franks, P., Jerant, A., Feldman, M., Duberstein, P., Fernandez y Garcia, E., Hinton, L., Strohecker, L., Kravitz, R.L., 2014. The effect of targeted and tailored patient depression engagement interventions on patient-physician discussion of suicidal thoughts: a randomized control trial. J. Gen. Intern. Med. 29 (8), 1148–1154. https://doi.org/10.1007/s11606-014-2843-8.
Siu, A.L., Bibbins-Domingo, K., Grossman, D.C., Baumann, L.C., Davidson, K.W., Ebell, M., Kemper, A.R., 2016. Screening for depression in adults: US preventive services task force recommendation statement. JAMA 315 (4), 380–387. https://doi.org/10.1001/jama.2015.18392.
Sobocki, P., Ekman, M., Agren, H., Krakau, I., Runeson, B., Martensson, B., Jonsson, B., 2007. Health-related quality of life measured with EQ-5D in patients treated for depression in primary care. Value Health 10 (2), 153–160. https://doi.org/10.1111/j.1524-4733.2006.00162.x.
Spitzer, R.L., Kroenke, K., Williams, J.B., Löwe, B., 2006. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch. Intern. Med. 166 (10), 1092–1097. https://doi.org/10.1001/archinte.166.10.1092.
Thombs, B.D., Ziegelstein, R.C., Roseman, M., Kloda, L.A., Ioannidis, J.P., 2014. There are no randomized controlled trials that support the United States preventive services task force guideline on screening for depression in primary care: a systematic review. BMC Med. 12, 13. https://doi.org/10.1186/1741-7015-12-13.
Torous, J., Nicholas, J., Larsen, M.E., Firth, J., Christensen, H., 2018. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid. Based Ment. Health 21 (3), 116–119. https://doi.org/10.1136/eb-2018-102891.
Trautmann, S., Beesdo-Baum, K., 2017. The treatment of depression in primary care: a cross-sectional epidemiological study. Dtsch. Arztebl. Int. 114 (43), 721. https://doi.org/10.3238/arztebl.2017.0721.
Vos, T., Lim, S.S., Abbafati, C., Abbas, K.M., Abbasi, M., Abbasifard, M., Murray, C.J.L., 2020. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the global burden of disease study 2019. Lancet 396 (10258), 1204–1222. https://doi.org/10.1016/S0140-6736(20)30925-9.

7.1.1. *Publication I: Supplementary Material*

Supplementary Fig. I: GETFEEDBACK.GP feedback.

Supplementary Fig. II: Standard (non-tailored) feedback.



Supplemental Fig. II. STANDARD FEEDBACK: All screens as displayed on the DISCOVER study website, including the layer 'How do I find an online therapy?' (English translation).

Supplementary Fig. III: Tailored feedback.

Supplemental Fig. III. TAILORED FEEDBACK: Examples for all four combinations of answers to the questions 'Do you think your symptoms are indications of depression?' and 'Do you worry about your symptoms?' as displayed on the DISCOVER study website (English translation).

7.2. Publication II

The efficacy of automated feedback after internet-based depression screening (DISCOVER): an observer-masked, three-armed, randomised controlled trial in Germany

Kohlmann, S., **Sikorski, F.**, König, H.-H., Schütt, M., Zapf, A., & Löwe, B.

Status: Published in The Lancet Digital Health, July, 2024.

Articles

The efficacy of automated feedback after internet-based depression screening (DISCOVER): an observer-masked, three-armed, randomised controlled trial in Germany

Sebastian Kohlmann, Franziska Sikorski, Hans-Helmut König, Marion Schütt, Antonia Zapf, Bernd Löwe

Lancet Digit Health 2024; 6: e446–57

Department of Psychosomatic Medicine and Psychotherapy (Prof S Kohlmann PhD, F Sikorski MSc, Prof B Löwe MD), **Department of Health Economics and Health Services Research** (Prof H-H König MD), **and Department of Medical Biometry and Epidemiology** (M Schütt MSc, Prof A Zapf PhD), **University Medical Centre Hamburg-Eppendorf, Hamburg, Germany; Department of General Internal Medicine and Psychosomatics, University Medical Centre Heidelberg, Heidelberg, Germany** (Prof S Kohlmann)

Correspondence to:
Prof Sebastian Kohlmann, Department of General Internal Medicine and Psychosomatics, University Medical Centre Heidelberg, Heidelberg 69120, Germany
sebastian.kohlmann@med.uni-heidelberg.de

Summary

Background Despite the availability of effective treatments, most depressive disorders remain undetected and untreated. Internet-based depression screening combined with automated feedback of screening results could reach people with depression and lead to evidence-based care. We aimed to test the efficacy of two versions of automated feedback after internet-based screening on depression severity compared with no feedback.

Methods DISCOVER was an observer-masked, three-armed, randomised controlled trial in Germany. We recruited individuals (aged ≥18 years) who were undiagnosed with depression and screened positive for depression on an internet-based self-report depression rating scale (Patient Health Questionnaire-9 [PHQ-9] ≥10 points). Participants were randomly assigned 1:1:1 to automatically receive no feedback, non-tailored feedback, or tailored feedback on the depression screening result. Randomisation was stratified by depression severity (moderate: PHQ-9 score 10–14 points; severe: PHQ-9 score ≥15 points). Participants could not be masked but were kept unaware of trial hypotheses to minimise expectancy bias. The non-tailored feedback included the depression screening result, a recommendation to seek professional diagnostic advice, and brief general information about depression and its treatment. The tailored feedback included the same basic information but individually framed according to the participants' symptom profiles, treatment preferences, causal symptom attributions, health insurance, and local residence. Research staff were masked to group allocation and outcome assessment as these were done using online questionnaires. The primary outcome was change in depression severity, defined as change in PHQ-9 score 6 months after random assignment. Analyses were conducted following the intention-to-treat principle for participants with at least one follow-up visit. This trial was registered at ClinicalTrials.gov, NCT04633096.

Findings Between Jan 12, 2021, and Jan 31, 2022, 4878 individuals completed the internet-based screening. Of these, 1178 (24%) screened positive for depression (mean age 37·1 [SD 14·2] years; 824 [70%] woman, 344 [29%] men, and 10 [1%] other gender identity). 6 months after random assignment, depression severity decreased by 3·4 PHQ-9 points in the no feedback group (95% CI 2·9–4·0; within-group d 0·67; 325 participants), by 3·5 points in the non-tailored feedback group (3·0–4·0; within-group d 0·74; 319 participants), and by 3·7 points in the tailored feedback group (3·2–4·3; within-group d 0·71; 321 participants), with no significant differences among the three groups (p=0·72). The number of participants seeking help for depression or initiating psychotherapy or antidepressant treatment did not differ among study groups. The results remained consistent when adjusted for fulfilling the DSM-5-based criteria for major depressive disorder or subjective belief of having a depressive disorder. Negative effects were reported by less than 1% of the total sample 6 months after random assignment.

Interpretation Automated feedback following internet-based depression screening did not reduce depression severity or prompt sufficient depression care in individuals previously undiagnosed with but affected by depression.

Funding German Research Foundation.

Introduction

Depressive disorders are among the most disabling and most prevalent disorders worldwide.[1] Despite the massive disease burden, depressive disorders often remain undetected and, therefore, untreated.[2] In turn, undetected depression increases the likelihood of a chronic course (ie, depression as a chronic condition), treatment resistance, increased health-care costs, and, most importantly, increased disease burden.[3] To prevent these far reaching consequences, interventions that can reach people who are affected by depression that is undetected at an early stage are needed.

Standardised depression screening could be one solution for early detection of those affected, but its effectiveness is debated due to weak empirical evidence.[4,5] Based on the recent US Service Preventive Task Force meta-analytic evidence report, it can be concluded that the effectiveness of depression screening

Research in context

Evidence before this study

Online depression screeners are promoted by mental health-care providers and are frequently used by people seeking diagnostic advice. We searched PubMed and MEDLINE for publications in English or German between June 1, 2023, and June 1, 1998 using the search terms "depression", "screening", "detection", "recognition", "internet-based", and "feedback". Additionally, we hand-searched reference lists of systematic reviews and meta-analyses. Both searches yielded similar results: treatment of depression and depression-related self-help after internet-based screening widely varied between 30% and 60% in observational studies. We identified one randomised controlled trial that concluded that feedback on screening results had no influence on depression severity. However, this trial was not sufficiently powered, had a high attrition rate that was associated with receiving the feedback intervention, did not exclude already diagnosed individuals, and had a short follow-up period of 3 months. Outside the internet setting, a randomised clinical trial in cardiology showed that written feedback of depression screening results to patients led to an improvement in depression severity and more information-seeking about depression after 6 months. In contrast, a three-arm randomised trial in primary care published in 2024 found no significant differences in depression severity after 6 months between study groups with different forms of written feedback after depression screening. However, there were some significant improvements in secondary clinical outcomes and healthcare uptake at 6 months in the group where patients received targeted feedback on their depression screening result compared with those who did not.

Added value of this study

To our knowledge, DISCOVER is the first adequately powered three-arm randomised controlled trial to test the efficacy of an internet-based screening and feedback procedure on depression severity. Moreover, it is the first trial that included only people who were affected by depression but had not yet been treated or diagnosed. The results of our trial indicated that automated feedback following internet-based depression screening did not reduce depression severity 6 months after random assignment compared with no feedback. Tailoring the feedback to individual participant characteristics and preferences did not affect this result. The initiation of evidence-based depression care or depression-related health behaviour was not associated with receiving feedback. The results did not change when adjusted for fulfilling the Diagnostic and Statistical Manual of Mental Disorders 5 (Text Revision)-based criteria for major depressive disorder or subjective belief of having a depressive disorder. Only one in seven participants reported having received a depression-related diagnosis by a health-care professional after 6 months and only one in four reported being treated for depressive disorder after 6 months.

Implications of all the available evidence

The DISCOVER randomised controlled trial suggests that automated feedback after internet-based depression screening does not change relevant depression outcomes. Even when internet-based screening combined with feedback refers screening positives to a health-care system that provides evidence-based depression care, a large proportion of those who screened positive did not receive adequate depression care. This finding should be considered by health-care providers offering automated feedback after internet-based depression screening and by guideline developers advocating depression screening for all adults.

depends on subsequent depression-related diagnostics and referral to evidence-based treatment.[4] An approach to stimulate this post-screening referral process is to provide feedback of screening results directly to the individual. Building on self-regulation theories of health behaviour,[6] it can be assumed that feedback of screening results could help individuals to recognise that they have depressive symptoms and motivate active functional health behaviour, such as seeking professional help. In line with these theoretical assumptions, a two-arm, single-centre, randomised controlled trial conducted in patients with coronary heart disease and hypertension suggested that feedback of depression screening results to patients and their treating physicians led to an improvement in depression severity and an increased search for information about depression 6 months after screening, compared with only feedback of screening results to the treating physicians.[7] Similarly, the results of a recently published three-armed multicentre randomised controlled trial in primary care suggest that feedback of depression screening results to previously undiagnosed patients can improve patient–physician communication and increase access to psychotherapy.[8] However, this multicentre trial did not find an effect of patient-feedback on depression severity 6 months after random assignment. Despite these results, subgroup analyses did suggest that feedback could lead to a reduction in depression severity in women, patients without addictions, and those with a history of lifetime depression. In both trials, there were no effects on depression severity 1 month after screening. Through the DISCOVER trial, we wanted to extend the evidence on feedback after depression screening to the internet.[9]

The internet can be regarded as one of the primary sources of information on mental health. According to user statistics from Mental Health America, the leading non-profit mental health organisation in the USA,

1 million people use the online depression screening tool on their website each year.[10] In Germany, one of the leading patient self-help organisations, Stiftung Deutsche Depressionshilfe, reports that half a million users complete the online depression screening tool on their website each year (Stiftung Deutsche Depressionshilfe und Suizidprävention, personal communication). Thus, internet-based depression screening combined with direct feedback could be a promising approach to reach people not yet reached by the health system. This screening and feedback procedure could guide individuals with depression to evidence-based care and have an effect on depression severity. In addition, the web-based technology also offers the possibility of tailoring the feedback to the characteristics of the individual. This approach is promising because, compared with non-tailored health information, tailored information is read more often, remembered better, and perceived as personally more relevant. Tailored health information can motivate people with depression to engage in treatment and help reduce the severity of depression.[11,12] Accordingly, tailored feedback of depression screening results could enhance the effect on non-tailored feedback.

The overall aim of the three-armed randomised controlled trial (DISCOVER) was to assess the efficacy of two different forms of automated feedback of screening results to individuals with at least moderate depression severity after internet-based depression screening. Additionally, DISCOVER aimed to investigate the effects of feedback on the initiation of evidence-based depression care and depression-related health behaviour, as well as potential negative effects of the screening and feedback intervention.

Methods

Study design and participants

We conducted an observer-masked, three-armed, randomised controlled trial to compare the change in depression severity in individuals given automated tailored feedback, automated non-tailored feedback, or no feedback of screening results 6 months after internet-based depression screening.

The study was promoted nationwide in Germany as a study on stress and psychological well-being. Traditional and social media campaigns (eg, advertisements on related websites, newsletters, Facebook, and Instagram) as well as print advertisements in public areas of several German cities (eg, flyers and posters) were used to approach interested individuals. To reach a sample that provided representativeness of the German population with respect to age and gender, a marketing company further advertised the study via a nationwide online access survey panel.

The study website was freely accessible and provided an internet-based eligibility check, baseline assessment, and depression screening. Individuals aged 18 years or older, who declared sufficient knowledge of German to answer questions on their health, provided contact details, and gave online informed consent could participate in an online eligibility assessment. The eligibility check was followed by the baseline assessment, including gender data via self-report. Participants then filled out an internet-based self-report depression rating scale (Patient Health Questionnaire-9; PHQ-9). Only individuals with at least moderate depression severity according to PHQ-9 (PHQ-9 ≥10 points) who indicated not being diagnosed or treated for depressive disorder within the previous year were included and randomly assigned automatically.

The Ethics Committee of the Hamburg Medical Chamber reviewed and approved the study (PV7039). The prospective trial registration and the statistical analysis plan are available at ClinicalTrials.gov (NCT04633096) and the detailed trial protocol has been published.[9] The trial was conducted and reported according to adequate CONSORT 2010 extensions and the CONSORT E-HEALTH statement.[13,14]

Randomisation and masking

After completing the depression screening, eligible participants were automatically and randomly assigned to receive no feedback, non-tailored, or tailored feedback on their screening results (1:1:1 allocation ratio). A computer-generated randomisation sequence was programmed by a trial-independent statistician and integrated in a computerised allocation system, which ensured allocation concealment of the research staff. The sequence consisted of permuted blocks of randomly arranged sizes (six, nine, and 12) and was stratified by baseline depression severity (moderate: Patient Health Questionnaire-9 [PHQ-9] score 10–14 points; severe: PHQ-9 score ≥15 points) to guarantee that the distribution among the study groups was the same across the depression severity levels. Individuals who participated multiple times in the internet-based eligibility check were automatically allocated to the same study group as before. This process was ensured by a privacy-preserving record linkage service that identifies double entries based on personal data and IP address.[15]

Outcome assessments were organised and conducted automatically using a computerised system. Research staff were masked to all outcome assessments as these were conducted via online questionnaires. Only the Structured Clinical Interview (SCID) for DSM-5 was conducted via telephone by research staff who were masked to allocation and instructed not to interact with the participants beyond the scope of the SCID. Masking of interviewers was checked by asking interviewers to document whether the participants disclosed their randomisation status and by asking the interviewers to guess the participant's study group. To assess the likelihood of agreement based on chance, intraclass κ values were calculated. Due to the design, the participants could not be masked but were kept unaware of trial hypotheses to minimise expectancy bias.

For more on the **study's promotion** see www.discover-studie.de

Articles

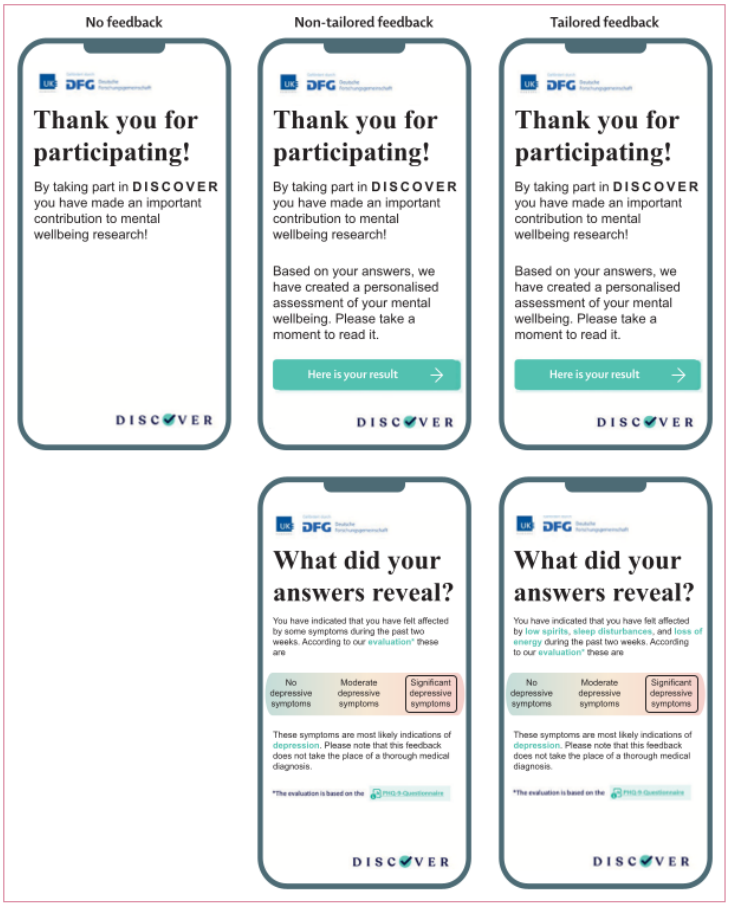


Figure 1: **Illustrations of the webpages for the no feedback group, the non-tailored feedback group, and the tailored feedback group**

Procedures

Directly after completing the internet-based depression screening, all participants received a thank you note and information on further follow-up assessments on their device screen. The participants who were randomly assigned to the no feedback group did not receive any feedback on the screening result. This group served as a passive control study group. The participants who were randomly assigned to one of the feedback groups were given the opportunity to receive feedback on their screening result by clicking on a link (figure 1).

See Online for appendix

The tailored and non-tailored feedback interventions both comprised four sections (appendix pp 2–6): the depression screening result (section 1), a note to seek diagnostic consultation by a mental health-care professional or general practitioner along with a hyperlink to make a direct appointment within the next 2 weeks (section 2), brief general information on depression (section 3), and information on depression treatment based on the German National Clinical Practice Guideline for Unipolar Depression (section 4).[16] Information was provided through direct hyperlinks to referenced health or social services (eg, online therapies and self-help groups) and the feedback form could be downloaded as a PDF that included all active links.

In the tailored feedback intervention (appendix pp 3–6), the information was adapted according to the participants' characteristics assessed as part of the baseline assessment. In section 1, the screening results were presented according to the participants' symptom profiles of the nine depressive symptoms assessed with the PHQ-9 (eg, you have indicated that you had low spirits, sleep disturbances, and loss of energy during the past 2 weeks). In section 2, the note to seek further diagnostic consultation was matched to the participants' specialist preferences (general practitioner *vs* mental health professional). In section 3, the information on depression was tailored to the participants' symptom profiles (eg, typical symptoms of depression are, for example, low spirits and sleep disturbances) and their symptom causal attributions (eg, triggers are, for example, stress with partner, negative thinking patterns, or a physical illness). In section 4, treatment options and help-seeking advice were adapted to the participant's health insurance providers (collected as part of the baseline assessment) and local residency in Germany (eg, by providing hyperlinks to self-help groups located nearby or to online therapies that are covered by the participant's health insurance provider). Additionally, after receiving the screening results (section 1) but before receiving further information (sections 2–4), the participants were asked via online self-report whether they thought that their symptoms were indications of depression and whether they worried about the symptoms. According to the participants' answers, the following three feedback sections were arranged automatically, phrased slightly differently, and extended by information tailored to the participant's risk profile (eg, depression can also occur during pregnancy).

The feedback interventions were developed in a multi-stage process together with those affected by depressive disorders.[17] For the current study, a digital graphic agency adapted the feedback material for presentation on a website. Throughout the process, the selection of content, design, and language was aligned to the current evidence on patients' needs in technology-based mental health interventions.[18] Responsive web design (ie, visualisation of the digital content automatically adjusted to the size of the screen) was applied to make study material user-friendly and reader-friendly for a range of devices (eg, smartphone, tablet, or laptop).

Outcomes

The primary outcome was the change in depression severity 6 months after randomisation with the PHQ-9. In accordance with the Diagnostic and Statistical Manual

of Mental Disorders 5 (Text Revision) diagnostic criteria, the PHQ-9 assesses nine depressive symptoms on a 0–3 scale, resulting in a total score ranging from 0 to 27, with higher scores representing higher depression severity.[19] A score of 10 or more points is recommended to screen for major depressive disorder. Based on a recent individual participant data meta-analysis of studies with a semi-structured interview reference standard, pooled PHQ-9 sensitivity is 0·85 (95% CI 0·79–0·89) and specificity is 0·85 (0·82–0·87).[20] The US Preventive Services Task Force and the German National Clinical Practice Guideline for Unipolar Depression recommend the PHQ-9 for depression screening.[4,16] To assess the differential effects of feedback in relation to the subjective belief of having a depressive disorder, the participants were asked whether they thought that they suffered from depression before random assignment. Additionally, participants were also interviewed via telephone 2–5 days after random assignment and 6 months after random assignment with the depression-related modules of the Structured Clinical Interview for DSM-5 (SCID). To ensure validity and reliability of the SCID, the interviewers (BSc or MSc Psychology) were trained and supervised by an experienced psychotherapist (PhD). Beyond the SCID interview instructions, the interviewers were instructed not to interact with the participants. Participants did not receive any feedback on the SCID results.

Secondary outcomes were the number of participants receiving evidence-based depression care (ie, psychotherapy or antidepressant treatment) since study start, the number of participants reporting to have received a diagnosis of a depressive disorder by a health-care professional since study start, the number of participants reporting to have engaged in depression-related health behaviour since study start (ie, information-seeking, self-management, seeking social support, and seeking formal help), health-related quality of life (EuroQoL-5 Dimension-5 Level visual analogue scale [EQ-5D-5 L VAS]),[21] anxiety severity (Generalized Anxiety Disorder Scale-7 [GAD-7]),[22] and somatic symptom severity (Somatic Symptom Scale-8 [SSS-8]).[23] To estimate possible unintended negative effects of the feedback interventions, the occurrence of any negative event attributed to trial participation was assessed 6 months after random assignment, with the following open question via telephone: in the last 6 months, have you experienced any negative changes that you attribute to participating in this study?

All outcome measures were self-reported by participants online. All outcomes were assessed at baseline and at the 6-month follow-up. In addition, participants were assessed with the PHQ-9 1 month after random assignment. The SCID interview was conducted 2–5 days after random assignment and again 6 months after random assignment. The SCID interview during the 6-month follow-up also included a question regarding negative effects attributed to trial participation. The SCID interviews were the only data assessments conducted by telephone.

To ensure participant safety, all participants who had indicated high suicidal ideation (PHQ-9 suicide item ≥2; almost every day) within the online questionnaire were automatically shown a webpage that provided advice to urgently seek help and relevant information on available help services (eg, general practitioner, local psychiatric emergency units, and the national emergency number; appendix p 8). If suicidal ideation was reported in one of the two interviews (2–5 days after random assignment and 6 months after random assignment), a mental health specialist from the study team was immediately consulted and contacted the patient by telephone. Patients with suicidal ideation who received consultations by a mental health specialist from the study team were excluded from the study, as this could have interfered with the feedback interventions.

Statistical analysis

The DISCOVER study was a superiority trial, testing pairwise comparisons between three study groups. According to the closed testing principle, the pairwise comparisons were only done if the overall *F* test of group (H0; all means are equal) could be rejected. All hypotheses were formulated as two-sided and tested for differences using a 5% significance level. The closed testing principle ensured a family-wise error level of 5%. The primary analysis was done as an ANCOVA of the PHQ-9 change scores (baseline to 6-month follow-up), including the factor group and the baseline value as a covariate, with subsequent pairwise comparisons of interventions by test of the corresponding contrasts to identify which groups are different. The PHQ-9 change score was included as the dependent variable using the original, continuous score. The assumptions regarding normality of residuals and linearity between the independent variables and the dependent variable were examined graphically. The primary analysis was done in the respective full analysis set population following the intention-to-treat principle, which included all participants who had been randomly assigned, regardless of whether they received the feedback or whether other protocol violations were known. For the full analysis set, at least a valid baseline and one valid post-baseline value of the primary outcome had to be available. For the primary analysis, only one timepoint was used and only those participants out of the full analysis set with a PHQ-9 score 6 months after randomisation were included in the primary analysis. In addition, in the case of missing follow-up values, we used the last observation carried forward principle as well as multiple imputations (using 100 generated datasets) as sensitivity analyses. As a further sensitivity analysis, we analysed data from the per-protocol population, which included participants who had no major protocol violations. A major protocol violation was predefined as having one of the following

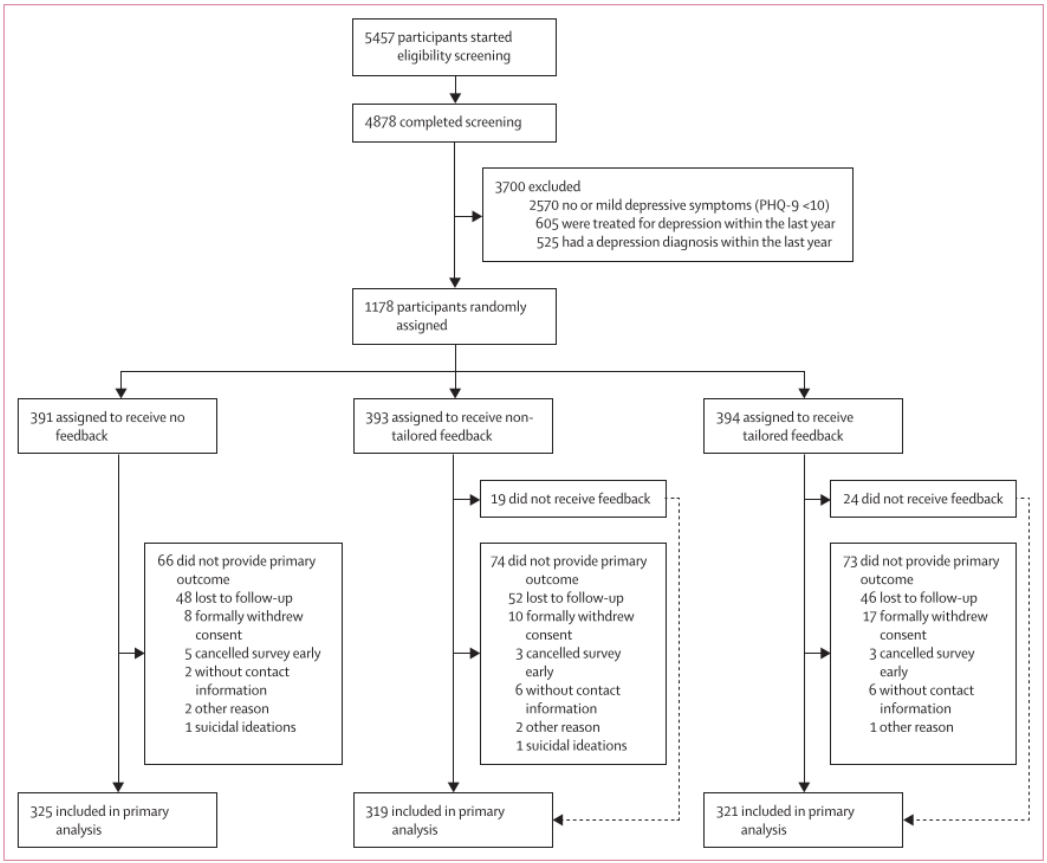


Figure 2: **Trial profile**
PHQ-9=Patient Health Questionnaire-9

deviations: not receiving the intervention (ie, feedback reading time <15 s or no download of feedback form), participating multiple times, indication at baseline that the survey was not seriously answered (ie, negative response to the question: have you answered the previous questions seriously?), completing the baseline survey in less than 2 min, or providing an invalid email address. We also conducted a prespecified subgroup analysis regarding the baseline depression severity (moderate: PHQ-9 ≥10 and ≤14 points; severe: PHQ-9 ≥15 points), in which the model of the primary analysis was extended by the binary factor depression severity (moderate *vs* severe) and the interaction between the group and the binary factor depression severity (moderate *vs* severe). A p value of the interaction of less than 0·05 would suggest that the depression severity at baseline was influencing the group effect. We calculated the within-group Cohen's d values using the respective estimated contrasts delivered by the primary analysis and the SDs were calculated using the raw data. Secondary outcome analyses were done in the full analysis set populations and were analysed for descriptive purposes. Secondary analyses were exploratory (ie, not hypothesis testing or generating evidence for efficacy). Thus, p values were not adjusted for multiple testing. For the secondary binary endpoints, the χ^2 test of independence was used to compare the three groups. For the continuous endpoints, ANCOVA was used as defined for the primary outcome. In addition, a multilevel model for the PHQ-9 change scores with participant as a random term was applied to the repeated measures in the same participant, including factor group and baseline value as covariates. In the secondary analyses, we repeated the analysis for PHQ-9 change 1 month after random assignment using all participants with a PHQ-9 score at that time.

All analyses were performed using SAS (9.4).

The study was powered to detect a small mean difference (Cohen's f 0·118) in the primary outcome (depression severity) in any pairwise comparison between all three study groups. The calculation was based on a global one-way ANCOVA adjusted for baseline depression severity, with an α of 0·05 (two-sided) and a

power of 80%. This resulted in a sample size of 233 participants per group. To allow for an estimated dropout of 35%, we aimed to recruit 358 participants per group.

Role of the funding source

The funder and the sponsor had no role in the study design, data collection, data analysis, data interpretation, or writing of the report.

Results

Recruitment took place between Jan 12, 2021, and Jan 31, 2022. The final follow-up assessment was on Sept 30, 2022. Of 5457 participants, 4878 (89%) participants completed the screening questionnaire (figure 2). Of those participants who completed the screening questionnaire, 2570 (53%) participants had no or mild depression severity (PHQ-9 <10), 605 (12%) were being treated for depression, and 525 (11%) had a depression diagnosis within the last year. In total, 1178 participants with at least moderate depression severity (PHQ-9 ≥10) were assigned to receive either no feedback (391 participants), non-tailored feedback (393 participants), or tailored feedback (394 participants) on the depression screening result. 6 months after randomisation, 965 participants provided follow-up data on the PHQ-9 (213 [18%] participants lost to follow-up).

Sample characteristics were well balanced across the three study groups (table 1). Across all three study groups, the mean participant age was 37·1 (SD 14·2) years, 824 (70%) participants were women, 344 (29%) were men, and ten (1%) reported having other gender identity, 121 (10%) had a migrant background (self-defined), 576 (49%) had a high education level, 484 (41%) were single, 847 (72%) were working (self-defined), and 600 (51%) indicated living in a large city. Individuals participated from across Germany, with most coming from the states of Hamburg, Rhine-Westphalia, Lower Saxony, and Bavaria. Most participants were recruited via social media, an online access panel, newsletter announcements, and search engines (appendix p 9). At baseline, the average PHQ-9 depression severity score was 14·8 (SD 4·0) and 165 (14%) participants reported that they currently did not think they currently had a depressive disorder. Of 909 participants interviewed with the SCID, 567 (62%) fulfilled the SCID criteria for major depressive disorder. The interviewers for the SCID did not know which group the participants had been assigned to, indicating that masking was successful (SCID interview conducted approximately 2–5 days after random assignment: κ −0·0202, p=0·52; and SCID interview 6 months after random assignment: κ −0·0237, p=0·48). With respect to the sample characteristics, there were no relevant differences between the population including all participants randomly assigned and the full analysis set population (appendix p 10).

	No feedback (n=391)	Non-tailored feedback (n=393)	Tailored feedback (n=394)
Age, years	36·5 (13·8)	37·7 (14·0)	37·2 (14·8)
Gender			
Woman	276 (71%)	275 (70%)	273 (69%)
Man	111 (28%)	115 (29%)	118 (30%)
Other gender identity	4 (1%)	3 (1%)	3 (1%)
German as first language	369 (94%)	370 (94%)	379 (96%)
Migrant background	47 (12%)	35 (9%)	39 (10%)
Being in a relationship	167 (43%)	200 (51%)	192 (49%)
Living with others	258 (66%)	277 (70%)	265 (67%)
Formal school education			
Low (<10 years)	71 (18%)	80 (20%)	68 (17%)
Middle (≥10 years)	120 (31%)	129 (33%)	134 (34%)
High (university entrance qualification)	200 (51%)	184 (47%)	192 (49%)
Working (self-defined)	276 (71%)	278 (71%)	293 (74%)
Quality of life (EuroQoL-5 Dimensions-5 Level visual analogue scale)	57·7 (22·4)	56·8 (22·2)	58·6 (22·0)
Depression severity (Patient Health Questionnaire-9)	14·8 (4·0)	14·8 (4·1)	14·7 (3·9)
Anxiety severity (Generalized Anxiety Disorder Scale-7)	12·0 (4·3)	12·3 (4·3)	11·9 (4·3)
Somatic symptom severity (Somatic Symptom Scale-8)	14·5 (5·3)	14·5 (5·1)	14·4 (5·3)
Number of depression-related risk factors*	6·0 (2·5)	6·1 (2·4)	5·8 (2·3)
Belief of having depressive disorder			
No	55 (14%)	44 (11%)	66 (17%)
Maybe	162 (41%)	201 (51%)	176 (45%)
Yes	174 (45%)	148 (38%)	152 (39%)
Fulfilling the Structured Clinical Interview for DSM-5 Disorders criteria for major depressive disorder	194 (62%)†	180 (61%)‡	180 (60%)§

Data are mean (SD) or n (%) unless stated otherwise. The Structured Clinical Interview for DSM-5 Disorders was conducted approximately 2–5 days after random assignment. *Risk factors included self-reported anxiety, addiction, traumatic life events, persistent physical symptoms, mood swings, chronic physical condition, lack of social support, mental comorbidity, mental comorbidity in family, history of suicide, current pregnancy, postnatal phase, menopause, and premenstrual syndrome. †78 cases with missing data. ‡97 cases with missing data. §94 cases with missing data.

Table 1: **Baseline and clinical characteristics of all randomly assigned participants (n=1178)**

374 (95%) of the 393 participants in the non-tailored feedback intervention opened the feedback screen and 370 (94%) of 394 participants in the tailored feedback intervention opened the feedback screen. 126 (34%) of the 374 participants who opened the feedback intervention in the non-tailored feedback group interacted with the feedback by clicking on at least one link or modal and 132 (36%) of the 370 participants in the tailored feedback group interacted with the feedback intervention. The median time spent on the feedback screen was 94 s in the non-tailored feedback group and 102 s in the tailored feedback group. 232 (59%) of 393 participants in the non-tailored feedback group and 229 (58%) of 394 participants in the tailored feedback group downloaded the PDF version of the feedback provided.

The primary analysis included 325 (83%) of 391 participants in the no feedback group, 319 (81%) of 393 participants in the non-tailored feedback group, and 321 (81%) of 394 participants in the tailored feedback group. The mean PHQ-9 score 6 months after depression

screening was 11·4 (SD 5·6) in the no feedback group, 11·2 (5·4) in the non-tailored feedback group, and 11·1 (5·5) in the tailored feedback group. Depression severity decreased by 3·4 (95% CI 2·9–4·0; within-group d 0·67) points on the PHQ-9 in the no feedback group, by 3·5 (3·0 to 4·0; within-group d 0·74) points in the non-tailored feedback group, and by 3·7 (3·2 to 4·3; within-group d 0·71) points in the tailored feedback group (figure 3). The overall *F* test did not indicate a difference in change in depression severity among study groups (p=0·72; table 2). The assumptions regarding normality of residuals and linearity between the independent variables and the dependent variable were met. Sensitivity analyses adjusting for missing values and protocol deviations did not indicate any differences among study groups and p values were 0·59 (full analysis set; multiple imputation using 100 generated datasets), 0·47 (full analysis set; last observation carried forward), and 0·50 (per-protocol dataset). The results did not change when adjusting for fulfilling the SCID-based diagnostic criteria for major depressive disorder or subjective belief of having depressive disorder. The subgroup analysis regarding the baseline depression severity showed no influence on the group effect (p=0·33).

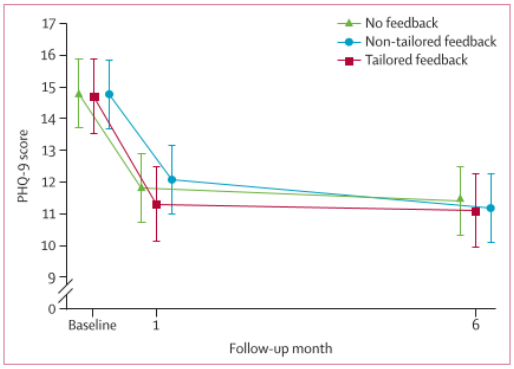


Figure 3: **Changes in depression severity over the course of 6 months of all randomly assigned participants (n=1178)**
Raw data plot of depression severity in the last 2 weeks as measured by the PHQ-9. Baseline depression severity was assessed before randomisation. Error bars show SDs. PHQ-9=Patient Health Questionnaire-9.

All secondary outcome analyses did not reveal relevant intervention effects (table 2). The analysis for the PHQ-9 change 1 month after random assignment showed no relevant differences between groups (no feedback group −3·0 [95% CI −3·4 to −2·5]; non-tailored feedback group −2·8 [−3·3 to −2·3]; tailored feedback group −3·5 [−4·0 to −3·0; appendix p 11). The multilevel model for the PHQ-9 change scores showed no relevant differences between groups. 6 months after random assignment, 43 (13%) of 324 participants in the no feedback group, 52 (16%) of 317 participants in the non-tailored feedback group, and 53 (17%) of 320 participants in the tailored feedback group reported having received the diagnosis of depressive disorder by a health-care professional within the last 6 months. Treatment initiation with antidepressants or psychotherapy was reported by 82 (25%) of 325 participants in the no feedback group, 93 (29%) of 319 participants in the non-tailored feedback group, and 91 (28%) of 321 participants in the tailored feedback group. The frequency of those seeking information on depression was similar between study groups, with 176 (55%) of

	No feedback	Non-tailored feedback	Tailored feedback	p value
Primary outcome				
Change in depression severity (Patient Health Questionnaire-9)	−3·4 (−4·0 to −2·9); 325	−3·5 (−4·0 to −3·0); 319	−3·7(−4·3 to −3·2); 321	0·72
Secondary outcomes				
Evidence-based depression care				
Diagnosis by a health-care professional	43 (13%); 324	52 (16%); 317	53 (17%); 320	0·43
Psychotherapy or antidepressant	82 (25%); 325	93 (29%); 319	91 (28%); 321	0·50
Depression-related health behaviour				
Seeking information*	176 (55%); 322	171 (54%); 317	180 (56%); 319	0·81
Seeking social support†	213 (66%); 325	201 (63%); 319	219 (68%); 321	0·38
Self-management‡	193 (59%); 325	197 (62%); 319	215 (67%); 321	0·12
Seeking formal help§	134 (41%); 325	143 (45%); 319	156 (49%); 321	0·17
Change in quality of life (EuroQoL-5 Dimensions-5 Level visual analogue scale)	4·1 (1·9 to 6·4); 321	3·9 (1·6 to 6·2); 312	3·2 (1·0 to 5·5); 318	0·85
Change in anxiety severity (Generalized Anxiety Disorder Scale-7)	−3·3 (−3·8 to −2·9); 323	−3·1 (−3·6 to −2·6); 314	−3·4 (−3·9 to −3·0); 318	0·65
Change in somatic symptom severity (Somatic Symptom Scale-8)	−2·6 (−3·2 to −2·1); 322	−2·2 (−2·8 to −1·7); 314	−2·5 (−3·1 to −2·0); 318	0·55

Data are in adjusted mean difference (95% CI), N, or n (%). Adjusted mean difference shows change from baseline to 6-months follow-up; adjusted for corresponding outcome at baseline. p values refer to *F* tests (continuous outcomes) or χ^2 tests (binary outcomes) and are not adjusted for multiple testing. *Seeking information included information obtained through personal contact, print media, or the internet. †Seeking social support included contact with peers, friends, family, or self-help groups. ‡Self-management included increasing physical activity, using relaxation techniques, improving sleep hygiene, or using unguided self-help programmes (books or internet-based). §Seeking formal help included seeking contact with primary care physicians or mental health professionals.

Table 2: **Primary and secondary outcomes 6 months after random assignment (full analysis set population)**

322 participants in the no feedback group, 171 (54%) of 317 participants in the non-tailored feedback group, and 180 (56%) of 319 participants in the tailored feedback group. The frequency of participants seeking social support for depression was also similar between groups, with 213 (66%) of 325 participants in the no feedback group, 201 (63%) of 319 participants in the non-tailored feedback group, and 219 (68%) of 321 participants in the tailored feedback group. 215 (67%) of 321 participants in the tailored feedback group engaged in depression self-management versus 197 (62%) of 319 participants in the non-tailored feedback group and 193 (59%) of 325 participants in the no feedback group. Compared with other depression-related health behaviour, seeking formal help for depression was reported least frequently across all study groups with 134 (41%) of 325 participants in the no feedback group, 143 (45%) of 319 participants in the non-tailored feedback group, and 156 (49%) of 321 participants in the tailored feedback group. Anxiety severity (GAD-7) and somatic symptom severity (SSS-8) decreased, and quality of life (EQ-5D-5 L VAS) increased 6 months after random assignment, with no difference among study groups being indicated. Results for the secondary analyses did not change when adjusting for fulfilling the SCID-based diagnostic criteria for major depressive disorder (appendix p 12).

Of 909 participants interviewed 6 months after random assignment, negative effects attributed to trial participation were reported by one participant in the no feedback group, by four participants in the non-tailored feedback group, and by four participants in the tailored feedback group. Trial participation was emotionally burdensome for one participant in the no feedback group, three participants in the non-tailored feedback group, and one participant in the tailored feedback group. One participant in the non-tailored feedback group and one participant in the tailored feedback group reported distressing memories associated with trial participation, and two participants in the tailored feedback group felt helpless after participating in the trial. 6 months after random assignment, suicidal ideation was reported during the telephone interview by one participant in the no feedback group and by one participant in the non-tailored feedback group. These participants received further consultations with a mental health specialist and acute suicidal ideation was ruled out. Suicidal ideation was not associated with study participation, as concluded by clinicians.

Discussion

The results of the DISCOVER trial show that automated feedback after internet-based depression screening did not reduce depression severity compared with no feedback. Contrary to our secondary hypothesis, tailoring the feedback to individual participant characteristics and preferences did not affect this result. Receiving feedback about the depression screening results was not associated with the initiation of evidence-based depression care or depression-related health behaviour. Irrespective of the study groups, only one in seven participants reported having received the diagnosis of depressive disorder by a health-care professional, and only one in four reported being treated for depression 6 months after internet-based depression screening. Negative effects of internet-based depression screening were reported by less than 1% of the total sample. Taken together, there are three main findings in our study. First, the primary result emphasises that even when automated feedback on depression screening results, tailored to a person's depression symptoms and treatment preferences, refers to a health system that covers depression treatment, it does not impact relevant depression outcomes. This finding should be considered by health-care providers offering internet-based depression screening and by guideline developers recommending that all adults should be screened for depression. Second, tailored feedback did not lead to greater interaction with the intervention compared with non-tailored feedback. Thus, future internet-based approaches aimed at the early detection of depression should weigh the effort of tailoring against the potential benefit. Third, there is no evidence from this trial that internet-based screening combined with automated feedback leads to relevant negative effects.

Despite the frequent use of internet-based depression screening globally, there is only one other trial that investigated effects on the uptake of mental healthcare.[24] Compared with the results of DISCOVER, this study showed that feedback after completing a lengthy internet-based assessment did not lead to increased uptake of mental healthcare when compared with general advice to seek help. However, this study had several methodological limitations, including a high attrition rate of 65% 3 months after random assignment, which was additionally associated with receiving feedback. The fact that a large proportion of participants in this trial[24] had already received support from mental health services could have influenced results on new uptake as well. In contrast, the DISCOVER trial was adequately powered and addressed individuals with undiagnosed depression who had not received care for depression before trial participation. To our knowledge, this is the first trial that provides methodologically sound and empirically robust evidence showing that automated feedback after depression screening does not improve relevant depression outcomes. This finding was unexpected as the feedback intervention referred to the German health-care system, where further diagnosis, evidence-based treatment, and appropriate follow-up are available and covered by social health insurance.

There are five potential explanations for this null finding. First, although almost two-thirds of participants downloaded the feedback form, only a third interacted

with the feedback on the internet (ie, clicked on hyperlinks). Contrary to our assumption, tailoring feedback according to individual participant characteristics and preferences did not lead to more engagement in depression care. Second, regarding the frequency of the feedback intervention, instead of one-time feedback, more prompts might be necessary to motivate those with depressive symptoms to seek evidence-based treatment. Third, the average PHQ-9 score was high at baseline but showed a clinically meaningful reduction 6 months after random assignment in all study groups. Even though the effect of the feedback intervention was expected to be small, this unexpectedly strong decrease in depression severity might have made it statistically difficult to detect a difference between the study groups. Fourth, regarding the frequent engagement in depression-related health behaviour, based on previous studies,[25,26] it was assumed that depression screening alone would not have a large effect on self-help behaviour. This ceiling effect might have worked against the primary hypothesis. Fifth, regarding the low uptake of evidence-based depression treatment, only a quarter of participants received evidence-based depression treatment. This finding highlights that, even when internet-based feedback referred to a health-care system that provides evidence-based depression care, many people who screened positive remained untreated.[27]

The results of this trial should also be considered in the context of its limitations. First, DISCOVER did not include a no screening study group. Therefore, the effect of screening alone cannot be inferred. Second, the study was announced as a survey on mental health and well-being but did not explicitly call for those seeking information on depression. This recruitment strategy was chosen as those affected are often unaware that their symptoms are indications of depression. However, those affected and actively seeking information on depression might have been more eager to follow the advice of the feedback intervention. Third, outcome data on help-seeking and health-care use were self-reported. Although this kind of assessment is common, additional data from health-care providers are necessary to estimate actual overuse, underuse, and misuse of health-care resources. Fourth, as the trial was internet-based, recruitment was self-selective, and study inclusion could not be verified in-person. However, this approach reflects the reality that the indication for internet depression screening is also self-selective. In addition, several mechanisms (eg, privacy preserving record linkage and telephone interviews) were applied to ensure fraud detection as well as the singularity and validity of the data. Fifth, assessment via SCID-interview and repeated assessments of depressive symptoms might have had an influence on depression course over the study period.

A major strength of the DISCOVER trial is its large sample size in combination with a lower-than-estimated loss-to-follow-up rate of 18% 6 months after random assignment. This facilitated an adequately powered analysis to detect even small effects. Furthermore, this trial is one of the first to disentangle the effect of screening and automated feedback from further depression care and is the first internet-based trial on early detection that included only participants affected by depression who were not treated or diagnosed. Regarding generalisability, the widespread advertisement across Germany and varying recruitment strategies should have ensured a sample that is representative of individuals interested in mental health. Finally, diagnostic interviews were done to adjust the efficacy analysis for the clinical diagnosis of depressive disorder.

The data of the DISCOVER trial highlight that internet-based depression screening can reach those who are affected by depression that is undetected. However, there is a gap between detection and referral to evidence-based depression treatment. Most individuals who screened positive initiated depression-related health behaviour, but few received evidence-based treatment. The primary analysis of this trial clearly shows that a feedback intervention after depression screening is not enough to bridge the gap between detection of depression and referral to evidence-based depression treatment. How to best facilitate access to depression care after internet-based screening is still not known. As internet-based treatment for depression is effective and widely accessible,[28] one solution could be to directly offer internet-based treatment to every individual who screens positive. However, this pragmatic approach undermines the necessity of a diagnostic consultation and accurate diagnosis as well as the importance of personal preferences and individual needs regarding depression treatment.[29] Combining artificial intelligence-based diagnostic techniques with personalised advice that accounts for individual treatment preferences when referring to evidence-based treatments could be an effective approach to initiating evidence-based depression care after screening.[30,31]

In conclusion, the results of the DISCOVER study indicate that automated feedback after internet-based screening neither reduces depression severity nor initiates evidence-based depression care. Health-care providers that offer internet-based depression screening should consider this finding. Furthermore, this result should inform the development of guidelines for early depression detection. Although DISCOVER indicates that there is no easy solution for addressing undetected depression, the results can stimulate future research to further understand the complex pathway from early detection to effective treatment.

Contributors

SK designed the study. BL, AZ, H-HK, and SK obtained the funding. FS and SK collected the data. AZ, MS, FS, and SK did the data analyses. SK wrote the original draft of the manuscript. All authors contributed to the interpretation of the data for the paper and critically reviewed and edited the original draft. All authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Declaration of interests

SK reports research funding (no personal honoraria) from the German Research Foundation and the German Federal Ministry of Education and Research. H-HK reports research funding (no personal honoraria) from the German Research Foundation, the German Federal Ministry of Education and Research, the German Innovation Committee at the Joint Federal Committee, and the European Commission Horizon 2020 Framework Programme. AZ reports research funding (no personal honoraria) from the German Research Foundation, the German Federal Ministry of Education and Research, the German Innovation Committee at the Joint Federal Committee, and the European Commission Horizon 2020 Framework Programme. BL reports research funding (no personal honoraria) from the German Research Foundation, the German Federal Ministry of Education and Research, the German Innovation Committee at the Joint Federal Committee, the European Commission Horizon 2020 Framework Programme, the European Joint Programme for Rare Diseases, the Ministry of Science, Research and Equality of the Free and Hanseatic City of Hamburg, and the Foundation Psychosomatics of Spinal Diseases, Stuttgart, Germany; has received remuneration for several scientific book articles from various book publishers and as a committee member from Aarhus University; has received travel expenses from the European Association of Psychosomatic Medicine and accommodation and meals from the Societatea de Medicina Biopsyhosociala for a presentation at the EAPM Academy at the Conferința Națională de Psihosomatică, Cluj-Napoca, Oct, 2023; and was a board member of the European Association of Psychosomatic Medicine (unpaid) until 2022. FS and MS declare no competing interests.

Data sharing

Individual participant data that underlie the results reported in this Article, after deidentification (text, tables, figures, and appendices), will be shared by the corresponding author on reasonable request for academic and research purposes and subject to data sharing agreements. The statistical analysis plan is available in the appendix (pp 13–29).

Acknowledgments

This work was funded by the German Research Foundation (424162019). We thank all participants who gave their consent to participate in DISCOVER and supported the study with data; Karl Wegscheider for his critical and constructive comments and his contribution to the quality of the manuscript; Anne Daubmann, who supported the biometric analyses and helped interpret the data; and Elisa Rabbiolo and Tarik Korkutan, who contributed to the organisation and recruitment of the trial.

References

1 GBD 2019 Mental Disorders Collaborators. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet Psychiatry* 2022; **9:** 137–50.

2 Vigo D, Haro JM, Hwang I, et al. Toward measuring effective treatment coverage: critical bottlenecks in quality- and user-adjusted coverage for major depressive disorder. *Psychol Med* 2020; **52:** 1–11.

3 Kraus C, Kadriu B, Lanzenberger R, Zarate CA Jr, Kasper S. Prognosis and improved outcomes in major depression: a review. *Transl Psychiatry* 2019; **9:** 127.

4 O'Connor EA, Perdue LA, Coppola EL, Henninger ML, Thomas RG, Gaynes BN. Depression and suicide risk screening: updated evidence report and systematic review for the US Preventive Services Task Force. *JAMA* 2023; **329:** 2068–85.

5 Thombs BD, Markham S, Rice DB, Ziegelstein RC. Does depression screening in primary care improve mental health outcomes? *BMJ* 2021; **374:** n1661.

6 Leventhal H, Phillips LA, Burns E. The Common-Sense Model of Self-Regulation (CSM): a dynamic framework for understanding illness self-management. *J Behav Med* 2016; **39:** 935–46.

7 Löwe B, Blankenberg S, Wegscheider K, et al. Depression screening with patient-targeted feedback in cardiology: DEPSCREEN-INFO randomised clinical trial. *Br J Psychiatry* 2017; **210:** 132–39.

8 Löwe B, Scherer M, Braunschneider L-E, et al. Clinical effectiveness of patient-targeted feedback following depression screening in general practice (GET.FEEDBACK.GP): an investigator-initiated, prospective, multicentre, three-arm, observer-blinded, randomised controlled trial in Germany. *Lancet Psychiatry* 2024; **11:** 262–73.

9 Sikorski F, König HH, Wegscheider K, Zapf A, Löwe B, Kohlmann S. The efficacy of automated feedback after internet-based depression screening: study protocol of the German, three-armed, randomised controlled trial DISCOVER. *Internet Interv* 2021; **25:** 100435.

10 Yom-Tov E, Lekkas D, Heinz MV, Nguyen T, Barr PJ, Jacobson NC. Digitally filling the access gap in mental health care: an investigation of the association between rurality and online engagement with validated self-report screens across the United States. *J Psychiatr Res* 2023; **157:** 112–18.

11 Levesque DA, Van Marter DF, Schneider RJ, et al. Randomized trial of a computer-tailored intervention for patients with depression. *Am J Health Promot* 2011; **26:** 77–89.

12 Shah R, Franks P, Jerant A, et al. The effect of targeted and tailored patient depression engagement interventions on patient-physician discussion of suicidal thoughts: a randomized control trial. *J Gen Intern Med* 2014; **29:** 1148–54.

13 Eysenbach G. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. *J Med Internet Res* 2011; **13:** e126.

14 Montgomery P, Grant S, Mayo-Wilson E, et al. Reporting randomised trials of social and psychological interventions: the CONSORT-SPI 2018 Extension. *Trials* 2018; **19:** 407.

15 Rohde F, Franke M, Sehili Z, Lablans M, Rahm E. Optimization of the Mainzelliste software for fast privacy-preserving record linkage. *J Transl Med* 2021; **19:** 33.

16 Härter M, Prien P, Group NVLG. Clinical practice guideline: the diagnosis and treatment of unipolar depression—national disease management guideline. *Dtsch Arztebl Int* 2023; **120:** 355–61.

17 Seeralan T, Härter M, Koschnitzke C, et al. Patient involvement in developing a patient-targeted feedback intervention after depression screening in primary care within the randomized controlled trial GET.FEEDBACK.GP. *Health Expect* 2021; **24** (suppl 1): 95–112.

18 Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. *Evid Based Ment Health* 2018; **21:** 116–19.

19 Kroenke K, Spitzer RL, Williams JB, Löwe B. The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: a systematic review. *Gen Hosp Psychiatry* 2010; **32:** 345–59.

20 Negeri ZF, Levis B, Sun Y, et al. Accuracy of the Patient Health Questionnaire-9 for screening to detect major depression: updated systematic review and individual participant data meta-analysis. *BMJ* 2021; **375:** n2183.

21 Günther OH, Roick C, Angermeyer MC, König HH. The responsiveness of EQ-5D utility scores in patients with depression: a comparison with instruments measuring quality of life, psychopathology and social functioning. *J Affect Disord* 2008; **105:** 81–91.

22 Löwe B, Decker O, Müller S, et al. Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. *Med Care* 2008; **46:** 266–74.

23 Gierk B, Kohlmann S, Kroenke K, et al. The somatic symptom scale-8 (SSS-8): a brief measure of somatic symptom burden. *JAMA Intern Med* 2014; **174:** 399–407.

24 Batterham PJ, Calear AL, Sunderland M, Carragher N, Brewer JL. Online screening and feedback to increase help-seeking for mental health problems: population-based randomised controlled trial. *BJPsych Open* 2016; **2:** 67–73.

25 BinDhim NF, Alanazi EM, Aljadhey H, et al. Does a mobile phone depression-screening app motivate mobile phone users with high depressive symptoms to seek a health care professional's help? *J Med Internet Res* 2016; **18:** e156.

26 Choi I, Milne DN, Deady M, Calvo RA, Harvey SB, Glozier N. Impact of mental health screening on promoting immediate online help-seeking: randomized trial comparing normative versus humor-driven feedback. *JMIR Ment Health* 2018; **5:** e26.

27 Trautmann S, Beesdo-Baum K, Knappe S, et al. The treatment of depression in primary care. *Dtsch Arztebl Int* 2017; **114:** 721–28.

28 Etzelmueller A, Vis C, Karyotaki E, et al. Effects of internet-based cognitive behavioral therapy in routine care for adults in treatment for depression and anxiety: systematic review and meta-analysis. *J Med Internet Res* 2020; **22:** e18100.

29 Karyotaki E, Efthimiou O, Miguel C, et al. Internet-based cognitive behavioral therapy for depression: a systematic review and individual patient data network meta-analysis. *JAMA Psychiatry* 2021; **78:** 361–71.

30 Carlbring P, Hadjistavropoulos H, Kleiboer A, Andersson G. A new era in internet interventions: the advent of Chat-GPT and AI-assisted therapist guidance. *Internet Interv* 2023; **32:** 100621.

31 Abd-Alrazaq A, Alhuwail D, Schneider J, et al. The performance of artificial intelligence-driven technologies in diagnosing mental disorders: an umbrella review. *NPJ Digit Med* 2022; **5:** 87.

7.2.1. Publication II: Supplementary Material

THE LANCET
Digital Health

Supplementary appendix

This appendix formed part of the original submission and has been peer reviewed. We post it as supplied by the authors.

Supplement to: Kohlmann S, Sikorski F, König H-H, Schütt M, Zapf A, Löwe B. The efficacy of automated feedback after internet-based depression screening (DISCOVER): an observer-masked, three-armed, randomised controlled trial in Germany. *Lancet Digit Health* 2024; **6:** e446–57.

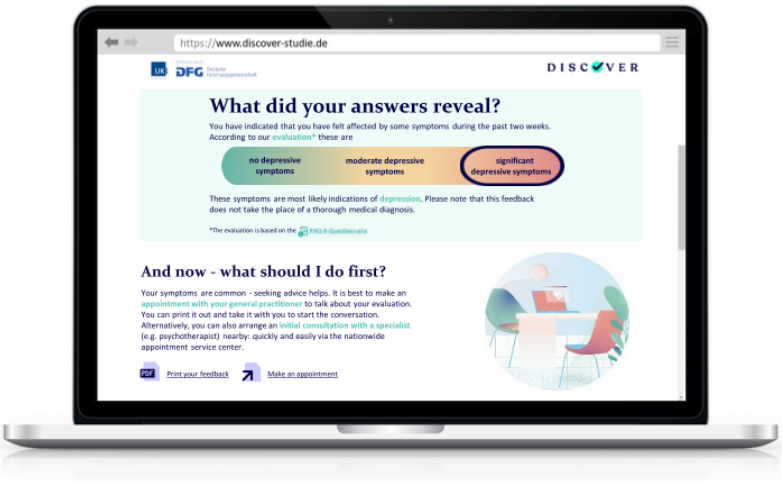
Supplementary material

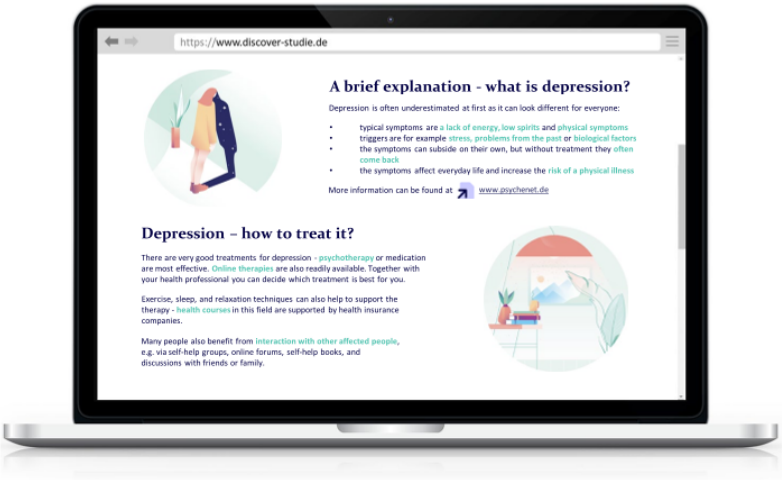
The efficacy of automated feedback after internet-based depression screening: The investigator-initiated, three-armed, randomised controlled trial DISCOVER

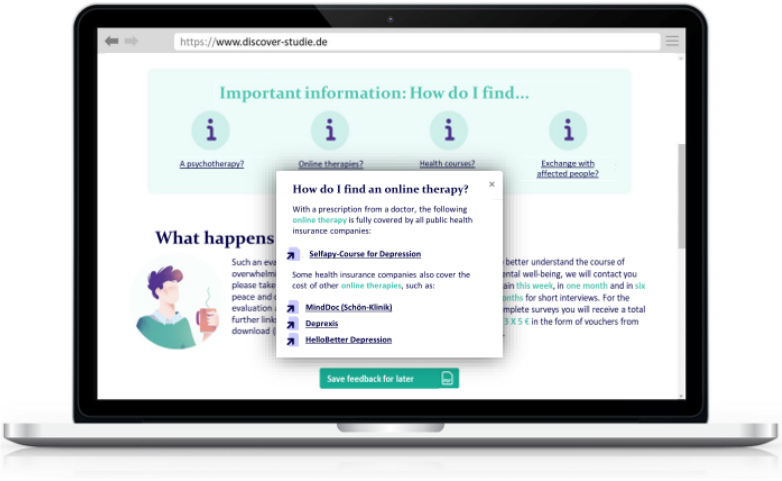
Sebastian Kohlmann et al.

Table of Contents

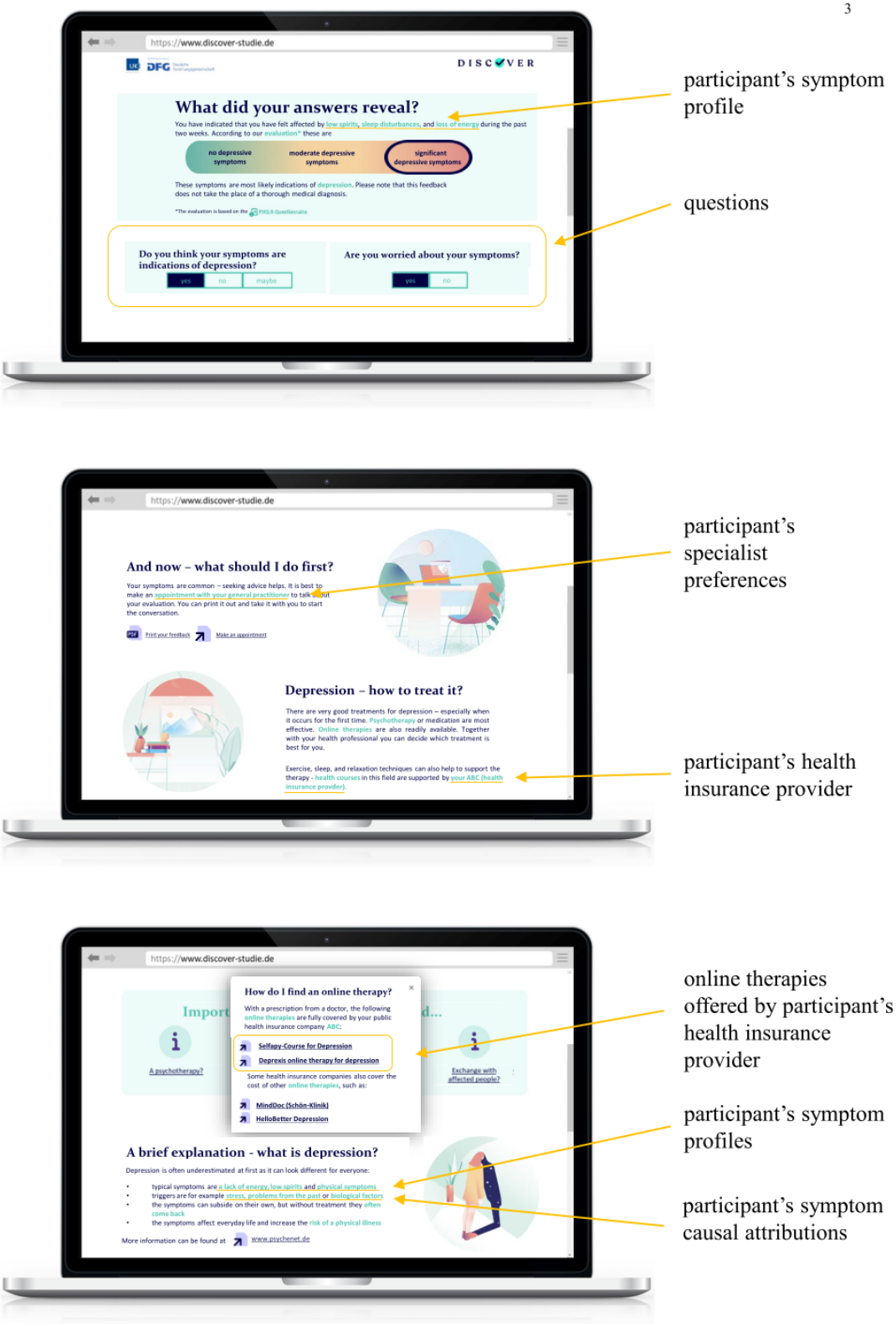
2



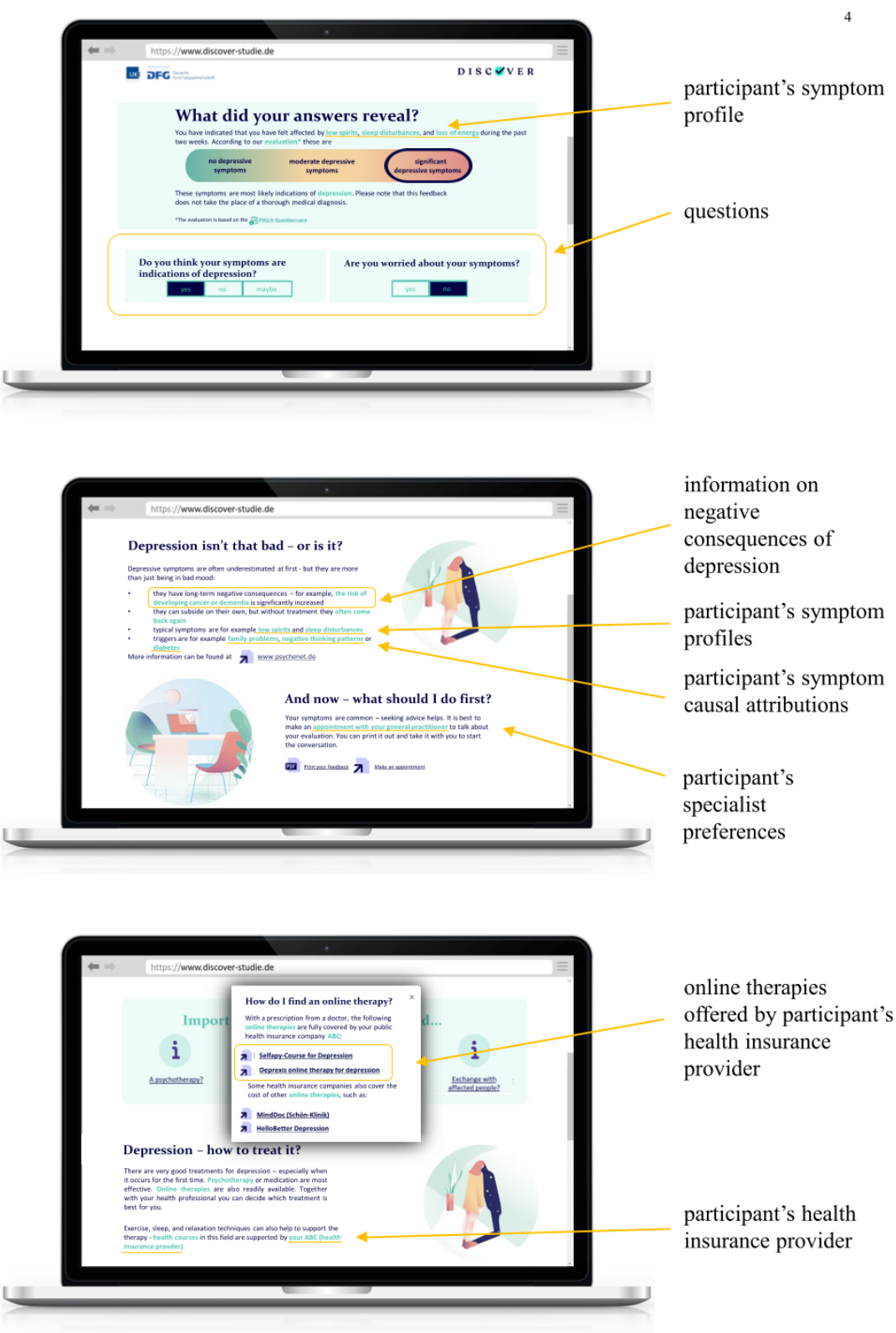




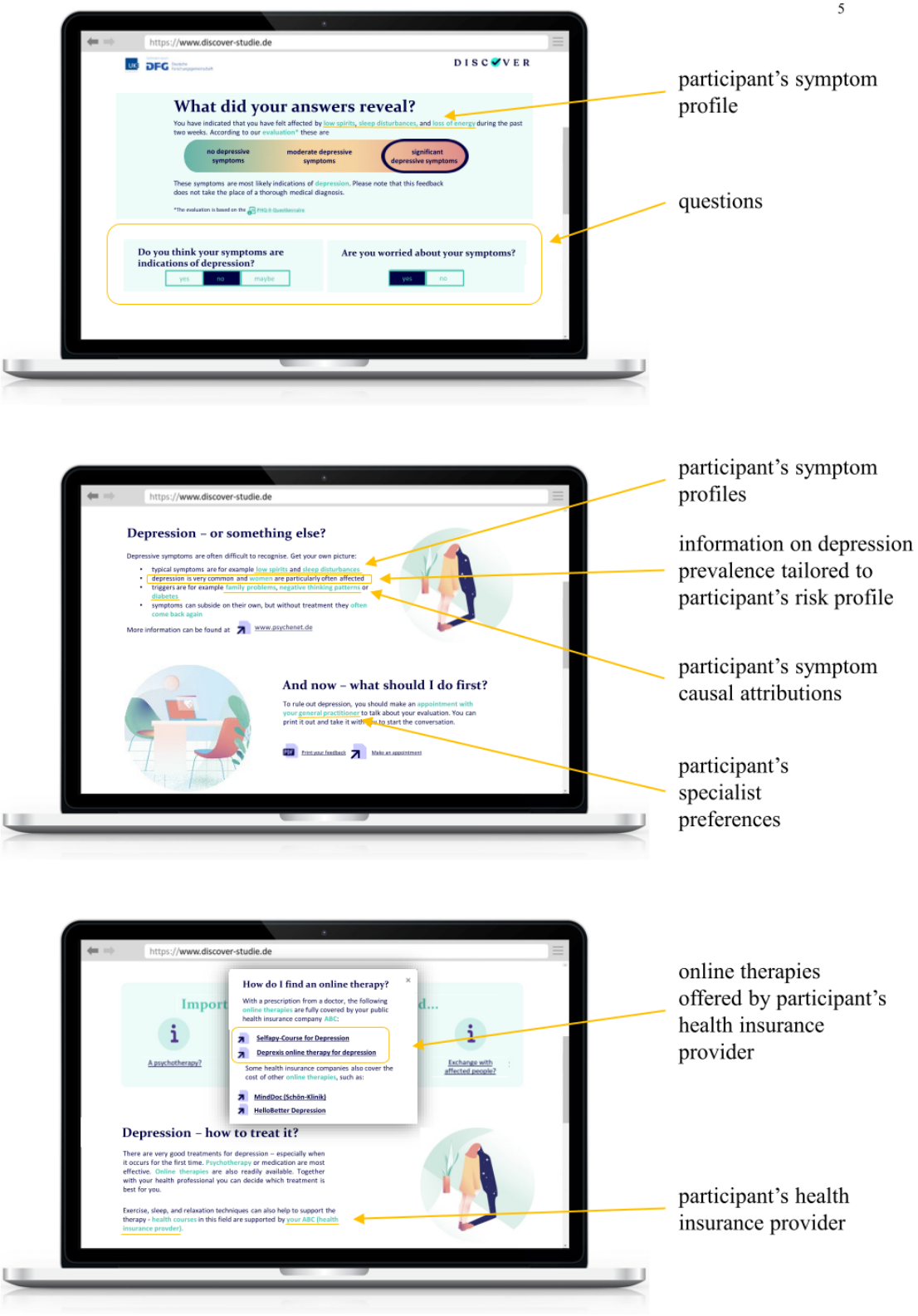
Supplemental Figure: Non-tailored feedback: All screens as displayed on the DISCOVER study website, including the layer ‘How do I find an online therapy?’ (English translation)



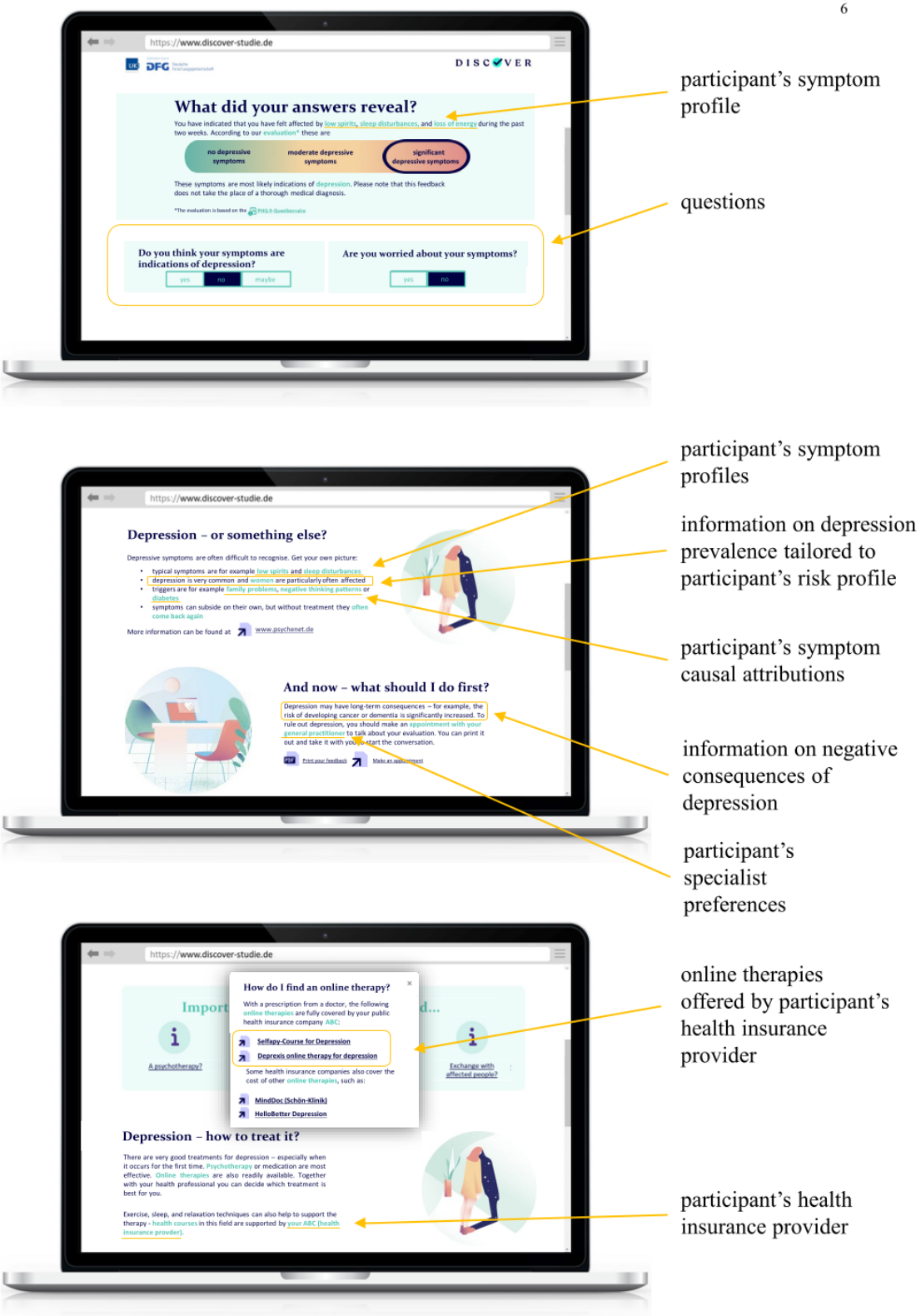
Supplemental Figure: Exemplary tailored feedback for participants who indicate that their symptoms are indications of depression and that they worry about them (English translation)



Supplemental Figure: Exemplary tailored feedback for participants who indicate that their symptoms are indications of depression but that they do not worry about them (English translation)

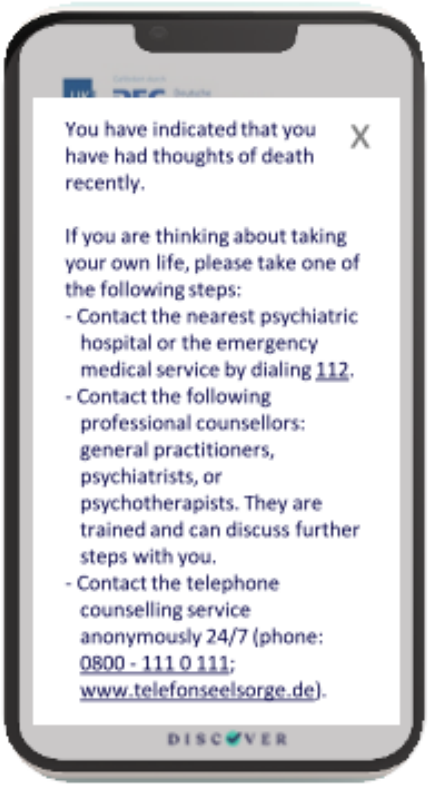


Supplemental Figure: Exemplary tailored feedback for participants who indicate that their symptoms are not indications of depression but that they worry about them (English translation)



Supplemental Figure: Exemplary tailored feedback for participants who indicate that their symptoms are not indications of depression and that they do not worry about them (English translation)

7



Supplemental Figure: Illustration of the emergency advice (mobile version) displayed to participants with elevated suicidal ideation (English translation)

https://arztsuche.116117.de/

Supplement 4 Information: Hyperlink to make a direct appointment health-care professional or general practitioner within the next 2 weeks provided by the Kassenärztliche Vereinigungen and the Kassenärztliche Bundesvereinigung on the medical on-call service in Germany.

	No Feedback (n=391)	Nontailored Feedback (n=393)	Tailored Feedback (n=394)
Personal recommendation	10 (3%)	14 (4%)	14 (4%)
Flyer	21 (5%)	18 (5%)	21 (6%)
Health insurance	26 (7%)	26 (7%)	24 (6%)
Newsletter	43 (11%)	30 (8%)	40 (10%)
Internet search	41 (10%)	40 (10%)	42 (11%)
Online-access panel	55 (14%)	55 (14%)	52 (13%)
Unknown	78 (20%)	97 (25%)	94 (24%)
Social media	113 (29%)	105 (27%)	102 (26%)

Data are n (%).

Supplement Table: **Recruitment sources of all randomised participants (n=1178)**

	No Feedback (n=325)	Nontailored Feedback (n=319)	Tailored Feedback (n=321)
Age, years	36·7 (13·7)	38·2 (13·8)	37·3 (14·5)
Gender			
Female	235 (72%)	225 (71%)	224 (70%)
Male	87 (27%)	91 (29%)	94 (29%)
Diverse	3 (1%)	3 (1%)	3 (1%)
German mother tongue	308 (95%)	300 (94%)	308 (96%)
Migration background	41 (13%)	31 (10%)	32 (10%)
Being in a relationship	145 (45%)	160 (50%)	152 (47%)
Living together	216 (67%)	220 (69%)	210 (65%)
Formal school education			
Low (less than 10 years)	57 (18%)	61 (19%)	50 (16%)
Middle (at least 10 years)	95 (29%)	101 (32%)	106 (33%)
High (A-level or above)	173 (53%)	157 (49%)	165 (51%)
Working	228 (70%)	231 (72%)	243 (76%)
Quality of life (EQ-5D-5L VAS)	57·4 (22·4)	56·9 (21·5)	58·4 (21·5)
Depression severity (PHQ-9)	14.8 (4.0)	14.7 (4.1)	14.8 (3.8)
Anxiety severity (GAD-7)	12.0 (4.3)	12.4 (4.3)	12.0 (4.3)
Somatic symptom severity (SSS-8)	14.5 (5.3)	14.5 (5.2)	14.3 (5.3)
No. of depression-related risk factors*	6·0 (2·6)	6·1 (2·5)	5·9 (2·3)
Belief of having depressive disorder			
No	48 (15%)	33 (10%)	55 (17%)
Maybe	130 (40%)	165 (52%)	142 (44%)
Yes	147 (45%)	121 (38%)	124 (39%)
Fulfilling the SCID criterion for major depressive disoder	182 (62%)§	169 (61%)#	186 (65%)+

Data are mean (SD) or n (%). PHQ-9=Patient Health Questionnaire-9. EQ-5D-5L=EuroQoL-5 Dimensions-5 Level scale. VAS=visual analogue scale. GAD-7=Generalized Anxiety Disorder-7. SSS-8=Somatic Symptom Scale. *Risk factors included self-reported anxiety, addiction, traumatic life events, persistent physical symptoms, mood swings, chronic physical condition, lack of social support, mental comorbidity, mental comorbidity in family, history of suicide, current pregnancy, postnatal phase, menopause, premenstrual syndrome. SCID=Structured Clinical Interview for DSM-5 Disorders; the interview was conducted approximately 2 to 5 days after randomisation. § 32 cases with missing data. # 40 cases with missing data. + 34 cases with missing data.

Supplement Table: **Baseline and clinical characteristics of the primary analysis full analysis set population**

	Adjusted mean difference (95% CI)* or n (%)						p value
	n	No Feedback	n	Nontailored Feedback	n	Tailored Feedback	
Primary outcome							
Change in depression severity (PHQ-9)	329	-3·0 (-3·4 to -2·5)	325	-2·8 (-3·3 to -2·3)	322	-3·5 (-4·0 to -3·0)	0·0933

P values refer to F test (continuous outcome) and is not adjusted for multiple testing. CI=Confidence Interval. PHQ-9=Patient Health Questionnaire-9. *Change from baseline to one-month follow-up; adjusted for PHQ-9 at baseline.

Supplemental Table: **Primary outcome one month after randomisation (all participants out of the full analysis set with one-month PHQ-9 value available)**

	Adjusted mean difference (95% CI)* or n (%)						p value
	n	No Feedback	n	Nontailored Feedback	n	Tailored Feedback	
Primary outcome							
Change in depression severity (PHQ-9)	294	-3·8 (-4·3 to -3·2)	279	-3·9 (-4·5 to -3·4)	287	-4·1 (-4·7 to -3·6)	0·6769
Secondary outcomes							
Evidence-based depression care							
Diagnosis by a health care professional	182	28 (15%)	167	32 (19%)	186	37 (20%)	0·4884
Psychotherapy and/or antidepressant	182	46 (25%)	169	62 (37%)	186	63 (34%)	0·0551
Depression-related health behaviour							
Seeking information+	181	112 (62%)	167	101 (60%)	186	123 (66%)	0·5138
Seeking social support§	182	122 (67%)	169	105 (62%)	186	136 (73%)	0·0854
Self-management#	182	110 (60%)	169	104 (62%)	186	129 (69%)	0·1533
Seeking formal help$	182	86 (47%)	169	87 (51%)	186	102 (55%)	0·3454
Change in quality of life (EQ-5D-5L VAS)	291	4·9 (2·5 to 7·2)	274	4·2 (1·8 to 6·6)	285	4·9 (2·5 to 7·3)	0·8915
Change in anxiety severity (GAD-7)	292	-3·7 (-4·2 to -3·2)	275	-3·5 (-4·0 to -3·0)	285	-3·8 (-4·3 to -3·3)	0·6515
Change in somatic symptom severity (SSS-8)	291	-2·9 (-3·5 to -2·3)	275	-2·4 (-3·0 to -1·8)	285	-2·9 (-3·5 to -2·3)	0·3789

P values refer to F tests (continuous outcomes) or Chi-Square-tests (binary outcomes) and are not adjusted for multiple testing. CI=Confidence Interval. PHQ-9=Patient Health Questionnaire-9. + seeking information included information obtained through personal contact, print media or the internet. § seeking social support included contact with peers, friends, family or self-help groups. # self-management included increasing physical activity, using relaxation techniques, improving sleep hygiene, or using unguided self-help programmes (books or internet-based). $ seeking formal help included seeking contact with primary care physicians or mental health professionals. EQ-5D-5L=EuroQoL-5 Dimensions-5 Level scale. VAS=Visual analogue scale. GAD-7=Generalized Anxiety Disorder-7. SSS-8=Somatic Symptom Scale-8. *Change from baseline to 6-months follow-up; adjusted for corresponding outcome at baseline.

Supplemental Table: **Primary and secondary outcomes six months after randomisation adjusted for fulfilling the Structured Clinical Interview for DSM-5 disorders criterion for major depressive disorder (full analysis set population)**

Statistical Analysis Plan (SAP)

for

Final Analysis

Full Study Title:	The Efficacy of Automated Feedback After Internet-based Depression Screening: the German, Three-armed, Randomised Controlled Trial DISCOVER
Short Study Title:	DISCOVER
EudraCT:	NCT04633096
Authors:	Marion Schütt
Study Protocol Version:	Published study protocol (Sikorski et al., 2021)
Sponsor's Protocol Code:	n.a.
Sponsor:	University Medical Center Hamburg-Eppendorf
Version / Date:	V01-0 30-09-2022
Effective Date:	Date of last signature

Responsible Bio statistician: Prof. Antonia Zapf University Medical Center Hamburg-Eppendorf, Department of Medical Biometry and Epidemiology	30.09.2022 Date	A. Zapf Signature
Responsible Employee / Representative of Responsible Employee : Anne Daubmann University Medical Center Hamburg-Eppendorf, Department of Medical Biometry and Epidemiology	30.09.2022 Date	Signature
Principal Investigator / Representative and Sponsor / Representative: Sebastian Kohlmann, PhD University Medical Center Hamburg-	30.09.2022 Date	S. Kohlmann Signature

Eppendorf, Department of Psychosomatic Medicine and Psychotherapy

Document History:

Version No.	Date	Chapter / Attachment	Reason for Review / Modification
0.1	29.06.2022	n.a.	n.a.
0.2	06.07.2022	All chapters	Comments project team
0.3	07.07.2022	All chapters	Comments project team
0.4	20.09.2022	All chapters	Comments project team
0.5	28.09.2022	All chapters	Review cycle in the project team
1.0	30.09.2022		Final

Abbreviations

Abbreviation	Definition
ANCOVA	Analysis of Covariance
Brief IPQ	Brief Illness Perception Questionnaire
CI	Confidence interval
CSSRI	Client Sociodemographic and Service Receipt Inventory
EFS	Evaluated for Safety Set
FAS	Full Analysis Set
GAD-7	Generalized Anxiety Disorder Scale (7 items)
IEC	Independent Ethics Committee
ITT	Intention-To-Treat principle
IRB	Institutional Review Board
LOCF	Last Observation Carried Forward
NMAR	Not Missing At Random
PD	Protocol Deviation
PHQ-9	Patient Health Questionnaire (Depression severity, 9 items)
PP	Per Protocol
SAP	Statistical Analysis Plan
SCID-5-CV	Structured Clinical Interview for DSM-5 Disorders
SSS-8	Somatic Symptom Scale (8 items)
USE	Usefulness Scale
VAS	Visual Analogue Scale

Content

1 Introduction

This Statistical Analysis Plan (SAP) is based on the published study protocol (Sikorski et al., 2021) and follows the guideline for statistical analysis plans (Gamble et al., 2017)
Some aspects of the statistical methods and the study design are already described in the study protocol. This SAP aims to further specify the procedures and statistical methods applied during the analysis of the study data.

The description of the health economic evaluation is not subject of this SAP.

1.1 Background and Rationale

Depression is one of the most disabling disorders worldwide, yet it often remains undetected. One promising approach to address both early detection and disease burden is depression screening followed by direct feedback to participants. Evidence suggests that individuals often seek information regarding mental health on the internet. Thus, internet-based screening with automated feedback has great potential to address individuals with undetected depression.

1.2 Study Objective

The study objective is to determine whether automated feedback after internet-based depression screening reduces depression severity as compared to no feedback.

Primary Hypothesis: The depression severity 6 months after screening is lower in at least one of the two feedback study arms (STANDARD FEEDBACK and/or TAILORED FEEDBACK) as compared to the NO FEEDBACK study arm.
Secondary Hypothesis: The depression severity 6 months after screening is lower in the TAILORED FEEDBACK arm as compared to the STANDARD FEEDBACK arm.

1.3 Study Endpoint(s)

1.3.1 Primary Endpoint

The primary endpoint of the study is the change in self-reported depression severity total score from baseline to six months after randomization, measured via the Patient Health Questionnaire-9 (PHQ-9; Kroenke et al., 2001; Löwe et al., 2004). The PHQ-9 consists of 9 items and is scored on a 4-point Likert scale (0–3), resulting in a total score ranging from 0 to 27 (0 = best, 27 = worst).

1.3.2 Secondary endpoints

- PHQ-9 total score change from baseline to one month after randomization.
- Depression diagnosis by a health care professional, resulting in a binary variable (yes / no), measured by self-report 6 months after randomization.
- Guideline-based depression care, using a self-developed questionnaire (psychotherapy and/or medication), based on the German National Clinical Practice Guideline for Unipolar Depression (DGPPN et al., 2015), resulting in a binary variable (yes / no), measured 6 months after randomization.

- Depression-related health behaviour, using a self-developed questionnaire (information-seeking, seeking social support, self-management, seeking formal help), resulting in a binary variable (yes / no), 6 months after randomization.
- Health-related quality of life will be assessed using the EuroQol-5D (EQ-5D-5L, Ludwig etal., 2018) six months after randomization, change from baseline. The EuroQol-5D index-value is generated using 5 dimensions: mobility, self-care, activity, pain and anxiety (each on a 5-point scale: 1 = best, 5 = worst). In addition, a visual analogue scale (VAS) is assessed.
- Anxiety severity measured via the 7-item Generalized Anxiety Disorder total score change (GAD-7, Spitzer et al., 2006, German version: Löwe et al., 2008) six months after randomization, change from baseline.
- Somatic symptom severity measured via the 8-item Somatic Symptom Scale total score (SSS-8, Gierk et al., 2014) six months after randomization, change from baseline.

1.3.3 Other endpoints

- Illness beliefs regarding depressive symptoms will be measured with a modified version of the well validated Brief Illness Perception questionnaire (Brief IPQ, Broadbent et al., 2006) and analysed per domain / item.
- Major depression diagnosis evaluated with the Depression related modules of the Structured Clinical Interview for DSM-5 Disorders (SCID-5-CV, Beesdo- Baum et al., 2019). Interviews will be conducted via telephone by a trained assessor.Sum score of number of critical life events: Two open questions assessing relevant positive and negative critical life events will be asked by phone 6 months after randomization and summed up.

1.3.4 Safety endpoints

To estimate possible unintended adverse events of the feedback intervention, participants are asked about the occurrence of any negative event that is attributed to the trial with an open question six months after randomization. Adverse events will be categorized by the investigator.

1.3.5 Participants characteristics

Participant characteristics recorded before randomization include

- sociodemographic data (e.g. age, gender, education, family status, rural/urban area living, local residency),
- medical data (diagnosis of and treatments for depression) and
- risk factors for depression onset (e.g. chronic somatic comorbidities, pregnancy, alcohol and nicotine consumption).

Domain	German question	Notes
Rural/urban area living	Wie groß ist die Stadt, in der Sie leben?	
	In welchem Bundesland leben Sie?	
	Wie sind Sie krankenversichert?	
Demographic data	Wie groß sind Sie (ggf. schätzen)?	
	Wie schwer sind Sie (ggf. schätzen)?	
Local residency	Was ist Ihre Muttersprache?	
	Was ist Ihre Staatsangehörigkeit?	
	Würden Sie sich selbst als Migrant/in bzw. Person mit Migrationshintergrund bezeichnen?	
Gender	Was ist Ihr Geschlecht?	
Age	Wie alt sind Sie?	

Family status	Was ist Ihr Familienstand?	
	Wie ist Ihre Wohn-/ Lebenssituation?	
Education	Was ist Ihr höchster Schulabschluss?	Summarized into 3 categories
	Was ist Ihr beruflicher Status?	
	Besteht im Moment oder bestand in den letzten sechs Monaten ein reguläres Arbeitsverhältnis (außer Minijob)?	
Nicotin consumption	Rauchen Sie (Zigaretten, E-Zigaretten, Zigarillos, Pfeife)?	
	Wie viele Zigaretten (bzw. E-Zigaretten, Zigarillos, Pfeifenköpfe, etc.) rauchen Sie am Tag?	
	Seit wie vielen Jahren rauchen Sie (wenn Sie weniger als 1 Jahr rauchen, geben Sie bitte 0 ein)?	
Alcohol consumption, AUDIT-C	Wie oft trinken Sie Alkohol? Wenn Sie an einem Tag Alkohol trinken, wie viele alkoholhaltige Getränke trinken Sie dann typischerweise? Wie oft haben Sie im letzten Jahr an einem Tag 6 oder mehr alkoholische Getränke getrunken?	sum score
Risk factor score	Leiden Sie unter Ängsten? Leiden Sie unter einer Sucht (Drogen, Computerspielsucht, Spielsucht, etc.)? Hatten Sie in der Vergangenheit ein Lebensereignis, das Sie bis heute belastet? Leiden Sie seit mindestens sechs Monaten unter anhaltenden körperlichen Beschwerden (z.B. Rückenschmerzen, Kopfschmerzen, Übelkeit)? Litten Sie im letzten Monat unter Stimmungsschwankungen oder gedrückter Stimmung? Haben Sie (eine) chronische körperliche Erkrankung(en) (z.B. Herzerkrankung, Diabetes, Asthma, etc.)? Fühlen Sie sich von Ihrem sozialen Umfeld unterstützt? (invert) Wurde bei Ihnen die Diagnose einer anderen psychischen/seelischen Erkrankung gestellt? Haben Sie Familienmitglieder, die unter psychischen Beschwerden leiden (z.B. Depressionen, Ängste, Essstörungen, Suchterkrankungen)? Gibt es in Ihrer Familiengeschichte bekannte Fälle von Suiziden oder Suizidversuchen? Versuchen Sie im Moment schwanger zu werden? Sind Sie momentan schwanger? Haben Sie in den letzten 6 Monaten entbunden? Stillen Sie momentan? Leiden Sie unter prämenstruellen Störungen (PMS)? Befinden Sie sich in Ihrer Menopause?	sum score
Somatic morbidity score	Um welche chronische(n) Erkrankung(en) handelt es sich? Herzerkrankung Diabetes Atemwegserkrankungen (z.B. Asthma, COPD, etc.) Darmerkrankungen Neurologische oder Nervenerkrankungen (z.B. Multiple Sklerose, Schlaganfall, etc.) Schmerzerkrankungen Krebs	sum score

	Rheumatologische Erkrankungen und/oder Gelenkserkrankungen sonstige Erkrankung(en)	
Depression history	Wurde bei Ihnen jemals die Diagnose Depression oder Burnout gestellt?	
	Wer hat die Diagnose erstmalig gestellt und Ihnen vermittelt?	
	Wann waren Sie das letzte Mal aufgrund von Depressionen oder Burnout in Behandlung?	
	Um welche Art der Behandlung handelt(e) es sich?	
	Können Sie sich vorstellen, aktuell an einer Depression zu leiden?	
Comorbidities	Wurde bei Ihnen die Diagnose einer anderen psychischen/seelischen Erkrankung gestellt?	
	Haben Sie (eine) chronische körperliche Erkrankung(en) (z.B. Herzerkrankung, Diabetes, Asthma, etc.)?	
Potential treatment preference	Wenn Sie vermuten würden, an einer Depression zu leiden, welche/n Behandler/in würden Sie als erstes aufsuchen?	

2 Study Methods

2.1 Trial Design

The DISCOVER trial is designed as an internet-based, observer-blinded, stratified randomized controlled clinical trial with three parallel groups (1:1:1), which is conducted nationwide in Germany.

Study arms:

- No feedback: The participants do not get any feedback on their screening result.
- Standard feedback: Participants receive standardized feedback comprising the following four sections 1) the depression screening result, 2) a note to seek diagnostic consultation by a health professional, 3) brief general information on depression and 4) information on depression treatment.
- Tailored feedback: Participants receive standard feedback tailored to their individual symptom profile, illness perceptions and preferences. Details regarding the tailored feedback arm are described in the published study protocol.

2.2 Randomization and Blinding

Randomisation is based on a computer-generated randomization sequence (1:1:1 allocation ratio), which was conducted by an independent researcher of the Department of Medical Biometry and Epidemiology and is not accessible to any other study team member. The sequence consists of permuted blocks of randomly arranged sizes (6, 9 and 12) and is stratified by baseline depression severity (moderate: PHQ-9 ≥ 10 and ≤ 14 points; severe: PHQ-9 ≥ 15 points). Allocation is performed by a computerized system, ensuring allocation concealment.
Individuals who participate multiple times are automatically allocated to the same study arm as before. This process is ensured by a privacy-preserving record linkage service which identifies double entries based on personal data.

Participants know their allocation due to the nature of the intervention but are kept unaware of trial hypotheses to minimise expectancy bias. The research staff assessing outcomes in the telephone interviews are blind to the allocation at any time.

2.3 Sample Size

Based on the results of the preceding DEPSCREEN-INFO trial (Löwe et al., 2016), the study is powered to detect a small mean difference (Cohen's $f = 0.118$) in the primary outcome (depression severity) in any pairwise comparison between all three study arms. The calculation is based on a global one-way ANCOVA adjusted for baseline depression severity, with an alpha of 0.05 (two-sided) and a power of 80%. It results in a needed sample size of $n = 233$ participants per group (PASS, 2008). To allow for an estimated drop out of 35% (c.f. Christensen et al., 2009), 358 participants per group are recruited (1074 in total).

2.4 Framework

The DISCOVER study is a superiority trial, testing pairwise comparisons between three study arms. All hypotheses are formulated two-sided and test for differences. To show superiority, we additionally look at the point estimate and the two-sided 95% confidence interval.

2.5 Statistical Interim Analyses and Stopping Guidance

No interim analysis is conducted.

2.6 Timing of Outcome Assessments

Table 1. Measures and assessment time points for endpoints

Measures	T0	T1	T2	T3
Primary outcome				
Depression severity, PHQ-9	x		x	x[a]
Secondary outcomes / other outcomes				
Depression diagnosis by a health care professional				x
Guideline-based depression care (psychotherapy and/or medication)				x
Depression-related health behaviour (e.g. information-seeking)				x
Anxiety severity, GAD-7	x			x
Somatic symptom severity, SSS-8	x			x
Health-related quality of life, EQ-5D-5L	x			x
Illness beliefs, Brief IPQ	x		x	x
Intervention adherence	x			
Critical life events				x[b]
Depression diagnosis, SCID		x[b]		x[b]
Adverse events				x[b]
Characteristics				
Sociodemographic data	x			
Medical data	x			
Risk factors for depression onset	x			

Note. T0 = before randomisation (The randomization takes place immediately after the T0 assessments are performed); T1 = 2 days after randomisation; T2 = 1-month follow-up, T3 = 6-months follow-up; PHQ-9 = Patient Health Questionnaire-9; EQ-5D-5L = EuroQol-5D 5-L; GAD-7 = Generalized Anxiety Disorder-7; SSS-8 = Somatic Symptom Scale-8; SCID = Structured Clinical Interview for DSM-5 Disorders; Brief IPQ = Brief Illness Perception Questionnaire.

[a] Primary outcome; [b] Measures assessed via telephone interview.

2.7 Timing of Final Analysis

The final analysis of the DISCOVER trial takes place as soon as the final visit of the last participant is completed, the data are collected, the queries are processed, and the database is locked. According to our current milestone plan, the final data transfer will take place early October 2022. This is followed by the final analysis.

3 Statistical Principles

3.1 Confidence Intervals and *P* Values

All applicable statistical tests are two-sided and are performed using a 5% significance level. All confidence intervals presented are 95% and two-sided. Analyses of secondary, other and safety outcomes are performed exploratory without adjustment for multiple testing.

3.2 Intervention Adherence and Protocol Deviations

Intervention Adherence
Adherence will be defined using feedback reading time ≥ 15 sec (yes / no) and/or download of feedback form.

Protocol deviations
Major protocol deviations are defined as follows:

- Participants that did not adhere to or did not receive the intervention (feedback reading time < 15 sec or no download of feedback form)
- Participants that have participated repeatedly for whatever reason
- Participants that report having participated in the study before
- Participants that report not having answered the survey seriously
- Participants with a baseline survey completion time < 2 min
- Participants with an invalid email address

3.3 Analysis Populations

Full Analysis Set (FAS)
The primary analysis is based on the full analysis set (FAS). It is as complete as possible and as close as possible to the Intention-To-Treat (ITT) principle which includes all randomized participants, as belonging to their randomization arm, regardless of whether they received the feedback or not, or whether other protocol violations are known. For the FAS at least a valid baseline and one valid post-baseline value of the primary outcome needs to be available.

Per Protocol Population (PP)
The Per Protocol population is a subset of the FAS and includes only participants who have no major protocol violation (see chapter 3.2).

Evaluated for Safety Set (EFS)
All randomized participants who were provided directly after the randomization with feedback or a 'thank you' note (no feedback) will be included into the Evaluated for Safety (EFS) set. Only participants who were reached by phone 6 months after randomization could be asked regarding possible adverse events.

4 Trial Population

4.1 Screening Data

Reporting of the number of screened participants (participants who started T0) and the number of screening failures.
Patients with an invalid email address will be regarded as not eligible after randomization. Due to technical reasons the validity could only be checked after randomization.

If available, baseline participant characteristics of the screening failures and reasons for exclusion are analysed using summary statistics.
Categorical data are summarized by absolute and relative frequencies. Continuous data are summarized by mean, standard deviation, median, first and third quartile, minimum, and maximum. These measures are presented for the total screening population.

4.2 Eligibility

The following eligibility criteria according to the study protocol are assessed within a self-report online survey at T0 directly before randomization:

Participants are required to
- be aged 18 years or above,
- show an indication for at least moderate depression (PHQ-9 ≥ 10 points),
- provide contact details,
- be willing to give informed consent.

Furthermore, participants who clicked on the website, gave informed consent and entered their personal data were seen as having
- sufficient German language proficiency,
- have internet access and
- have sufficient computer/internet literacy.

Participants are excluded
- if they were diagnosed with depression within the past 12 months or
- if they currently are or were receiving depression treatment within the past 12 months.

Further, in some cases technical circumstances occurred before randomization and impeded randomization.
Due to technical reasons the validity of the email address could only be checked after randomization. Participants with an invalid email address will be regarded as not eligible after randomization.

4.3 Recruitment / Withdrawal / Follow-up

A CONSORT flow diagram is used to summarize the number of participants who were:

- assessed for eligibility at screening (started T0)
- eligible at screening and completed T0
- ineligible at screening*
- eligible and randomized
- eligible but not randomized*
- received the randomized feedback arm or no feedback

- did not receive the randomized feedback / no feedback
- lost to follow-up*
- discontinued the intervention*
- randomized and included in the primary analysis
- randomized and excluded from the primary analysis*

*reasons are provided

4.4 Baseline Participant Characteristics

Summary statistics are given for the baseline participant characteristics (see chapter 1.3).
Categorical data are summarized by absolute and relative frequencies. Continuous data are summarized by mean, standard deviation, median, inter-quartile range, minimum, and maximum. These measures are presented for the total screening population and by study group.

5 Analysis

5.1 Outcome Definitions

Primary Outcome

- PHQ-9 total score (9 items) change (T3 (6 months after randomization) – T0 (baseline))

Secondary Outcome

The following secondary endpoints are continuous outcomes. For these, difference from baseline (T0) is calculated for 1 month (T2; PHQ-9 only) and 6 months after randomization (T3, all outcomes):

- PHQ-9 total score (9 items)
- GAD-7 total score (7 items)
- SSS-8 total score (8 items)
- EQ-5D-5L (index-value, VAS)

Other Outcomes

The following other endpoints are continuous outcomes:

- Brief IPQ single items (difference from baseline (T0) is calculated for 1 month (T2) and 6 months after randomization (T3))
- Critical life events at T3 (sum score for positive and negative life events, respectively)

The following other endpoints are binary outcomes. For these, no difference to baseline (T0) is calculated:

- SCID at T1 and T3

Safety Outcome

- Adverse events at T3

5.2 Missing Data

The multilevel modelling approach limits the bias when handling missing data even in the case of not missing at random (NMAR).
As a sensitivity analysis missing values are imputed by different approaches for the primary analysis. In the FAS population no baseline value is missing as per definition. In case of missing follow-up values,

we conduct last observation carried forward (LOCF) and multiple imputation. In the LOCF approach, missing values are replaced by the previous valid value. The imputation model in the multiple imputation approach includes all baseline characteristics of participants and all variables of the primary outcome analysis. From the baseline values of the secondary outcomes, we include as many variables in the imputation model as possible. For the imputation of missing follow-up values, 100 imputed data sets are generated, and the results are combined using Rubin's Rules.

5.3 Efficacy Evaluation

Analysis of Primary Endpoint

A multilevel model incorporating the participants as a random term is applied to the repeated measures in the same participant, including the factor group and the baseline value as a covariate. The primary analysis is performed within the framework of this model as an ANCOVA of the PHQ-9 change scores (T0 to T3-difference), with subsequent pairwise comparisons of interventions by test of the corresponding contrasts. The primary analysis is performed in the respective FAS population. The pairwise comparisons are done if the global test of group (H_0: all means are equal) could be rejected. Each test will be performed at a two-sided level of alpha = 0.05. This closed testing principle will ensure a family-wise error level of 5%.

Absolute, and relative frequencies of missing observations are presented both overall and separately for the randomized groups.

Analysis of Secondary Endpoints

The secondary endpoint analyses are performed in the respective FAS population. All secondary endpoints are compared and statistically assessed for descriptive purposes and not in a confirmatory sense. The aim of the analysis is an exploratory data analysis, not hypothesis testing or generation of evidence for the efficacy. No attempt is made to adjust for the p-values for multiple testing.
Mean, standard deviation, first and third quartile, minimum and maximum for the continuous variables and absolute and relative frequencies for the categorical and binary variables are presented both overall and separately for the randomized groups for each time point. Number of missing observations are presented for the randomized groups separately for each time point. All summary tables are structured with a column for each group (tailored feedback, standard feedback, no feedback) and a total column.
Multilevel linear models with differences from baseline as outcome, incorporating the participants as a random term, are applied to the repeated measures in the same participant, including the factor group, time point, and the baseline value as a covariate. The interaction between time and the group is determined. If it is not significant, it is eliminated from the model. Adjusted means with 95% confidence intervals (CI) and p-values are reported.

Analysis of other Endpoints

Other endpoints will be analysed, depending on the level of scale, as defined for the secondary endpoints. See chapter 5.3. The SCID will be analysed descriptively using summary statistics.

Sensitivity Analyses

The primary endpoint analysis is repeated within the PP population.

Further for the primary endpoint a multilevel linear model with change from baseline as outcome incorporating the participants as a random term are applied to the repeated measures in the same

participant, including the factor group, time point, and the baseline value as a covariate. Adjusted means with 95% CI and p-values are reported.
See section 5.2 for sensitivity analyses regarding missing value imputation.

In addition, a subgroup analysis regarding the baseline depression severity (moderate: PHQ-9 ≥ 10 and ≤ 14 points; severe: PHQ-9 ≥ 15 points) is planned. The model defined for the primary analysis of the PHQ-9 change score will be extended by the factor depression severity (moderate / severe). The interaction of the feedback-group and the depression severity will be investigated. In case the p-value of the interaction is < 0.05 we consider that the depression severity at baseline is influencing the PHQ-9 effect.

5.4 Safety Evaluation

All adverse events will be analysed in the EFS using summary statistics and, if applicable, analysed using logistic regression.

5.5 Additional Analyses

Check of Assumptions

Normality of residuals and linearity
The assumptions regarding normality of residuals, and linearity between the independent variables and the dependent variable are examined graphically. Partial residual plots are used to examine linearity. Residual plots and quantile-quantile plots are used to evaluate normality of residuals. In case of unmet assumptions, the Box-Cox transformation is applied in order to find out which power transformation is reasonable in terms of not violating the assumption.

Drop Out Analysis

If applicable, we perform a drop out analysis. Therefore, we use a binary variable (drop out: yes vs. no) as a dependent variable in a logistic regression. The baseline characteristics as well as the baseline values of primary and secondary outcomes are used as effects.

Control for Blindness of the Participants

Steps to control for blindness include the following: after every interview, assessors are (a) instructed to document if participants have disclosed their randomisation status and (b) asked to guess the study arm. After study closure, this guess is compared with the actual status and Cohen's kappa is computed to identify whether hit rates differ from what can be expected from chance.

5.6 Data Challenges

Due to the internet format different scenarios can threaten data validity. Therefore, before the analysis a pre-processing of the data was performed to eliminate invalid entries. Pre-processing was performed by the statisticians, if applicable, and by the principal investigator, if checking of personal data was required. Repeated entries of the same participant were excluded. In case of randomized participants only the first entry was used. In case of not randomized participants the entry with the most information was used.

5.7 Differences to Trial Protocol

In contrast to originally stated in the study protocol, the USE will not be analysed. A change of the USE questionnaires occurred during the study.

Furthermore, website use will not be analyzed as a variable on its own, but will be used for the evaluation of major protocol deviations and intervention adherence (see 3.2).

5.8 Statistical Software

- SAS© 9.4 or newer

6 References

- Sikorski F, König H-H, Wegscheider K, Zapf A, Löwe B, Kohlmann S. The efficacy of automated feedback after internet-based depression screening: Study protocol of the German, three-armed, randomised controlled trial DISCOVER. Internet Interventions 25. 2021; 100435.
- International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) Guideline E9. Note for guidance on statistical principles for clinical trials (CPMP/ICH/363/96).
- CONSORT Statement (Consolidated Standards of Reporting Trials)
- Gamble C, Krishan A, Stocken D, Lewis S, Juszczak E, Doré C, et al. Guidelines for the Content of Statistical Analysis Plans in Clinical Trials. JAMA. 2017;318(23):2337-43

7.3. Publication III

Does feedback after internet-based depression screening cause harm? A secondary analysis of negative effects in the randomised controlled DISCOVER trial

Sikorski, F., Löwe, B., Daubmann,, A., & Kohlmann, S.

Status: Under review in Journal of Medical Internet Research.

Original Paper

Does feedback after internet-based depression screening cause harm? A secondary analysis of negative effects in the randomised controlled DISCOVER trial

Authors: Franziska Sikorski, Bernd Löwe, Anne Daubmann, & Sebastian Kohlmann

Affiliations:

Department of Psychosomatic Medicine and Psychotherapy, University Medical Center Hamburg-Eppendorf, Hamburg, Germany (F Sikorski MSc, Prof B Löwe)

Department of Medical Biometry and Epidemiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany (A Daubmann MSc)

Department of General Internal Medicine and Psychosomatics, University Medical Centre Heidelberg, Heidelberg, Germany (Prof S Kohlmann)

Corresponding author:

Franziska Sikorski MSc

Department of Psychosomatic Medicine and Psychotherapy, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

Martinistr. 52

20251 Hamburg

Germany

E-mail: franziska.sikorski@gmail.com

ORCID: 0000-0003-1191-5073

Abstract (494 words)

Background: Internet-based depression screening is frequently used and promoted by mental health providers. Recent evidence, however, indicates that it does not create substantial health benefit and there is criticism that it might be associated with unnecessary harm. Yet, systematic empirical evidence on postulated negative effects is missing.

Objective: We aimed to examine whether automated feedback after internet-based depression screening is associated with misdiagnosis, mistreatment, deterioration in depressive symptoms, deterioration in emotional response to symptoms, and deterioration in suicidal ideation.

Methods: This secondary analysis entails data from the German-wide, randomised controlled DISCOVER trial testing the efficacy of automated feedback after internet-based depression depression. Undiagnosed but affected individuals with a positive depression screening score (Patient Health Questionnaire-9, PHQ-9 ≥ 10 points) were randomised 1:1:1 to receive no, non-tailored feedback, or tailored feedback on their screening result together with recommendations to seek professional diagnostic advice. Participants were followed-up at one and six months via online questionnaires. Misdiagnosis and mistreatment were operationalised as having received a depression diagnosis by a health professional and as having started psychotherapy or antidepressant medication since screening while not meeting the DSM-5 criteria of a major depression (diagnostic telephone interviews). Deterioration in depressive symptoms was defined as a pre-post change of ≥ 4.4 points in the PHQ-9, deterioration in emotional response to symptoms as a pre-post change of ≥ 3.1 points in a composite scale based on the Brief Illness Perception Questionnaire, and deterioration in suicidal ideation as a pre-post change of ≥ 1 point in the PHQ-9 suicide item. Outcome rates were compared between both feedback arms and the no feedback arm in terms of relative risks (RR). Data collection was conducted between Jan 12, 2021, and Sept 30, 2022.

Results: In the per protocol sample of 948 participants (72% female; mean [SD] age, 37.3 [14.1] years), rates of misdiagnosis (six months; 3.5%-4.9%), mistreatment (six months; 7.2%-8.3%), deterioration in depression severity (one and six months; 2%-6.8%), deterioration in emotional response (one and six months; 0.7%-2.9%), and deterioration in suicidal ideation (six months; 6.8%-13.1) were not higher in the feedback arms compared to the no feedback arm (RRs ranging between 0.46 and 1.96, ps ≥ 0.13). The rate for deterioration in suicidal ideation at one month was higher in the non-tailored feedback arm (RR=1.92; p = 0.01), but not in the tailored feedback arm (RR=1.26, P = 0.43), with rates of 12.3%, 8.1%, and 6.4% in the non-tailored, the tailored, and the no feedback arm,

respectively. All but one sensitivity analyses supported the general findings and there were no indications for differential effects in the subgroup of false positive screens.

Conclusions: The results indicate that feedback after internet-based depression screening is not associated with negative effects such as misdiagnosis, mistreatment, and deterioration in depression severity or in emotional response to symptoms. However, it cannot be ruled out that non-tailored feedback may increase the risk of experiencing deterioration in suicidal ideation. Robust prospective research on negative effects and particularly suicidal ideation is needed and should inform current practice of public internet-based depression screening.

Registrations:

Trial Registration: ClinicalTrials.gov (NCT04633096)
Preregistration of data analysis: OSF.io (https://osf.io/tzyrd)

Keywords: internet-based, depression screening, early detection, negative effects, harms

Introduction

Depressive disorders, although being among the most disabling and most prevalent disorders worldwide [1], often remain undetected and therefore untreated [2]. In the last decades, depression screening has been increasingly discussed as promising to reach those affected but undetected at an early stage: In addition to population level screening in routine clinical care, as for example recommended in the United States [3], advocates also speak out in favour of screening for depression online [4]. For many affected individuals, the internet is already the favoured source for information on mental health [5, 6]. Further, so called internet-based depression tests are widely promoted by mental health-related institutions and frequently used by those seeking diagnostic advice [7]. The rational of internet-based depression tests typically involves administering symptom-based screening questionnaires and providing individuals with direct feedback on screening results, sometimes supplemented by links or referrals to services. The feedback is thought to empower affected individuals to better act on their symptoms [8] and to seek diagnostic consultation and, if necessary, appropriate care. As such, it might improve early detection and management of depression.

However, feedback after internet-based depression screening has been proposed and implemented without due consideration of its appropriateness, i.e. without evaluating its effectiveness against the background of potential negative effects. While assessing negative effects is recommended for all clinical interventions, it is particularly important in screening interventions, as in these per definition a substantial amount of participants will not benefit [9]. Additionally, in internet-based contexts there is no experienced health care stuff available to monitor participants who might need support [10]. As such, the balance between harms and benefits of feedback after internet-based depression screening could easily lean towards harms. Indeed, there is no empirical evidence of positive effects on targeted patient-related outcomes: two randomised controlled trials, one by Batterham and colleagues [11] and one recently conducted by our research team [12], do not indicate that feedback of internet-based depression screening results promotes the uptake of evidence-based depression care or reduces depression severity. Negative effects, if present, would therefore likely be generated without creating substantial health benefits.

Evidence in this regard is, however, scarce, with the current scientific debate being mainly reflected by opinion papers: The first area of negative effects of depression screening, discussed in both medical and internet-based contexts, relates to inadequate management and care for individuals who receive false positive feedback. Critics particularly point to the risk

of increased rates of misdiagnosis and mistreatment following screening. This, again, is assumed to lead to unnecessary iatrogenic effects such as adverse medication and psychotherapy side effects in healthy individuals, societal costs, and waste of limited health care resources resulting in potential undertreatment of more severe cases [4, 13, 14]. A second area of concern relates to negative psychological effects to the feedback of screening results. It is assumed that the associated labeling, resembling a clinical diagnosis, might induce anxiety, distress, stigma, or nocebo effects such as for example deterioration of symptoms [4, 14, 15]. These effects could be amplified by the fact that, in contrast to medical settings, in internet-based depression screening the 'diagnosis' would be delivered without a health professional who could provide emotional support or advice on further steps [16]. Indeed, in qualitative studies on internet-based mental health screening some participants describe having been discouraged, shocked or concerned by the feedback they received [8, 17]. Further, one observational study found that internet-based screening procedures including referrals to in-person care had a higher likelihood of subsequent online searches for suicidal intent, potentially suggesting a deterioration of suicidal ideation [18]. In contrast, in our recently conducted trial on feedback after internet-based depression screening less than 1% of participants reported on adverse events attributed to trial participation when qualitatively asked six months after screening [12]. However, robust and large-scale quantitative research on the postulated potential negative effects is missing.

Objective

In the present study, we addressed this lack of evidence by analysing data from our recently conducted randomised controlled trial DISCOVER that tested the efficacy of feedback after internet-based depression screening in a population of undiagnosed but affected individuals [12]. Based on the potential negative effects discussed in the literature and on outcomes assessed in that trial, we aimed to examine whether feedback after internet-based depression screening is associated with increased misdiagnosis, mistreatment, deterioration in depressive symptoms, deterioration in emotional response to symptoms, and deterioration in suicidal ideation one and six months after the screening.

Methods

The DISCOVER trial [19] and this secondary analysis [20] were pre-registered. We conducted small deviations from the preregistration: we added the outcomes misdiagnosis and

emotional response to symptoms, as we deemed this of clinical interest. Further, we added logistic regressions as sensitivity analyses to test for the robustness of findings. The detailed study protocol [17] and main outcomes [12] of the trial have been described previously. Online informed consent via checkboxes was obtained from all participants. The study was approved by the Ethics Committee of the Hamburg Medical Chamber (# PV7039) and followed appropriate Consolidated Standards of Reporting Trials (CONSORT) guidelines, including the harms and the e-health statement [9, 21-25] (see Multimedia Appendix 1). Data collection was conducted online and in German language between January 12, 2021, and September 30, 2022; this secondary data analysis was conducted between May 3, 2023, and December 23, 2023.

Study design and participants

DISCOVER was an investigator-initiated, observer-blinded, three-armed, randomised controlled trial that compared automated feedback with no feedback after internet-based depression screening. After being screened for depression with the digitised Patient Health Questionnaire-9 (PHQ-9 [26]), eligible participants were randomised to receive either no feedback, non-tailored feedback or tailored feedback on their screening result (1:1:1 allocation ratio). Assessments were set at baseline, one-month, and six-months follow-up. In this secondary analysis, we compared rates of misdiagnosis and mistreatment (at six months) as well as deterioration in depression severity, deterioration in emotional response to symptoms, and deterioration in suicidal ideation (at one and six-months) between each feedback arm and the no feedback arm.

Participants were individuals aged 18 years or above with at least moderate depression severity (PHQ-9 $\geq$ 10) but not diagnosed with or treated for depression within the last year. Additional eligibility criteria were having sufficient internet literacy and German language proficiency, providing contact details, and giving online informed consent.

Study procedures

The study was promoted as being on ‘stress and psychological well-being’ on a publicly accessible study website [27] from January 2021 to January 2022. The aim of evaluating internet-based depression screening was not explicitly communicated, but interested participants were informed that some of them will get feedback on a part of their answers. Traditional and social media campaigns as well as print advertisements in public areas of several German cities were used to approach interested individuals. To reach a sample that strives for representativeness of the German population with respect to age and gender, a

marketing company further advertised the study via a nationwide internet-based access survey panel.

After completing baseline assessment and screening, eligible participants were automatically randomised by random permuted blocks randomisation stratified for baseline depression severity (moderate: PHQ-9 score 10 to 14 points; severe: PHQ-9 score ≥ 15 points) and allocated 1:1:1 to one of the three study arms. Double entries identified based on personal data by a privacy-preserving record linkage service [28] were automatically re-allocated to their former study arm. Research staff were masked to allocation at any time until breaking the blind. Due to the design, participants could not be masked but were kept unaware of trial hypotheses to minimise expectancy bias.

Internet-based follow-up assessments were set at one month and six months after randomisation. Two to five days and six months after randomisation, participants were contacted via telephone for complementary diagnostic interviews (see [16] for more detailed information on data collection procedures). Participants were compensated for participation in follow-up assessments with up to 15 euros.

Internet-based depression screening and feedback of screening results

Participants underwent depression screening as part of the baseline survey using the digitised PHQ-9 [26, 29] (see the outcomes section for further information and Multimedia Appendix 2 for the layout of the digitised version). At the standard cut-off value of ≥10 points, the paper-pencil PHQ-9 demonstrates high discriminatory performance for detecting major depression with sensitivity ranging between 0.79 and 0.89 and specificity ranging between 0.82 to 0.87 [30]. Preliminary evidence suggests that psychometric characteristics are comparable for the digitised version [31, 32]. The PHQ-9 is recommended for depression screening by the US Preventive Services Task Force and the German National Clinical Practice Guideline for Unipolar Depression [3, 33].

After completing the baseline survey, all participants were thanked for participating in the study. Participants of the feedback arms received information on follow-up procedures and were offered feedback on their screening result by clicking on a ‘next’-button (Figure 1). Both non-tailored and tailored feedback comprised four sections: (1) the depression screening result, (2) a note to seek diagnostic consultation by a health professional together with a link to make an appointment within the next two weeks, (3) brief general information on depression, and (4) information on depression treatment based on the German National Clinical Practice Guideline for Unipolar Depression [33]. Notably, in the German health care system depression care is available and covered by the social health insurance. Information

was extended by direct links to referenced health or social services (e.g. internet-based therapies covered by the health insurance, self-help groups), and the feedback form could be downloaded in a file that included all active links. In extension to the non-tailored feedback, the information in the tailored feedback intervention was personalised to participants' characteristics (e.g., 'You have indicated that you had l*ow spirits, sleep disturbances,* and *loss of energy* during the past two weeks.'). Additionally, after being provided with the screening result (section 1) but before receiving further information (sections 2 to 4), participants were asked whether they think that their symptoms were indications of depression and whether they worried about the symptoms. According to the participants' answers, the following three feedback sections were arranged in a differing order, phrased slightly differently, and extended by information tailored to participants' risk profile (e.g. 'Depression *in pregnancy* is common.'). The feedback was developed in a multistage process involving patient representatives [34, 35] and a digital graphic agency to adapt the material to the possibilities of internet-based presentation. Illustrations of the complete non-tailored and tailored feedback versions can be found in Multimedia Appendices 3 and 4.

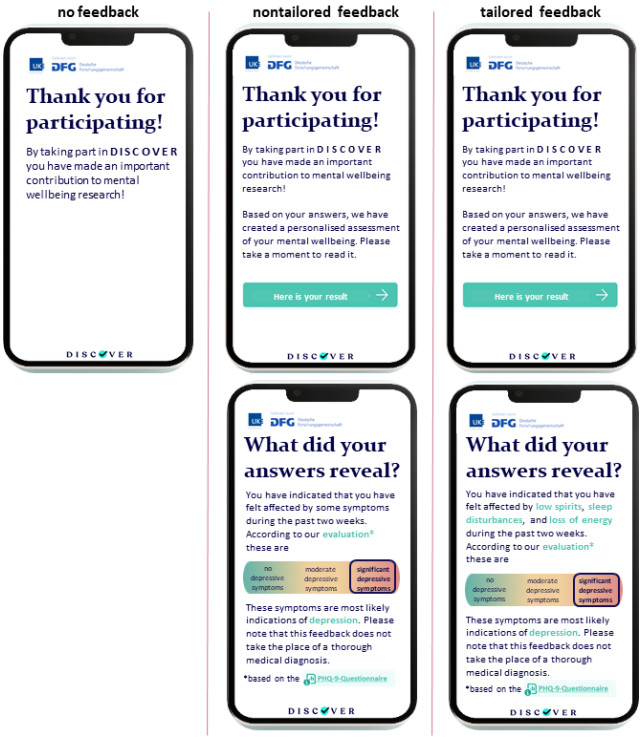


Figure 1. Illustrations of no feedback, non-tailored feedback (first screen), and tailored feedback (first screen) (reprinted from [12]).

Due to ethical considerations, all participants who have indicated elevated suicidal ideation (PHQ-9 suicide item ≥ 2; *more than half the days*) were shown a screen providing an advice to urgently seek help and relevant information on available help services (e.g. general practitioner, local psychiatric emergency units, and the national emergency number; Multimedia Appendix 5).

Measures

Depression diagnosis by a health professional was assessed at six months with the question: "Have you been diagnosed with depression or burnout in the last six months?". Guideline-based depression treatment, i.e. pharmacotherapy with antidepressant medication and/or psychotherapy as recommended by the German National Clinical Practice Guideline for Unipolar Depression [33], was assessed at six months with the questions: "Have you started taking medication to treat depression or other complaints such as sleep problems, anxiety or stress [which ones]?" and "Have you started any psychotherapy or similar treatment in the last 6 months [which]?". Participants could choose from guideline-based treatment options or give open answers. In case of open answers, these were checked for guideline-conformity independently by two of the authors (SK and FS).

Criteria for major depression at baseline were assessed with the depression-related modules of the Structured Clinical Interviews for DSM-5 Disorders (SCID-5-CV) [36] two to five days after screening. The interviewers (BSc / MSc psychology) were trained and supervised an experienced psychotherapist. Participants who did not meet the criteria for a major depression were considered as false positive screens.

Depression severity was assessed with the PHQ-9 at one and six months after screening. In accordance with the DSM-5 diagnostic criteria, the PHQ-9 assesses nine depressive symptoms each rated in terms of frequency during the past two weeks (0–3; *not at all* to *nearly every day*), resulting in a total score ranging from 0 to 27, with a higher score indicating higher depressive symptoms. The PHQ-9 is among the most frequently used self-report depression questionnaires, has good psychometric properties, and is sensitive to change [26, 37].

Suicidal ideation was assessed with the PHQ-9 suicide item (item 9): "Over the last two weeks, how often have you been bothered by thoughts that you would be better off dead or of hurting yourself in some way?", rated from 0–3 (*not at all, several days, more than half the days, nearly every day*).

Emotional response to depressive symptoms was assessed with a composite scale based on two items of the Brief Illness Perception Questionnaire (Brief IPQ) that cover

emotional representations of depressive symptoms: “How concerned are you about your symptoms?” and ”How much do your symptoms affect you emotionally (e.g. do they make you angry, scared, upset or depressed)?”. The items were assessed directly after the PHQ-9 and were scored on a Likert scale ranging from 0 (*not at all*) to 10 (*extremely*). Item scores were pooled for the composite scale, resulting in a total scale ranging from 0 to 10. The respective items of the Brief IPQ showed good psychometric properties [38].

Outcomes

Participants were classified as misdiagnosed or mistreated if they reported having received a depression diagnosis by a health professional or having started psychotherapy or antidepressant medication while not having met the criteria for major depression at baseline (SCID), i.e. while being screened false positive.

Deterioration in depression severity was defined as a pre-post change score of at least 4.4 points in the PHQ-9. The cut-off is based on the reliable change index (RCI), a psychometric criterion to evaluate whether a change in symptoms is considered statistically reliable, i.e. not attribuTable to measurement error [39]. The present RCI was calculated using the PHQ-9 standard deviation (SD) from the current sample ($SD_{baseline}$ = 4), the reliability coefficient from the PHQ-9 validation study (r_{tt} = 0.84) [40], and a 95% confidence level. The resulting RCI of 4.4 points is comparable to cut offs found in prior research [41, 42].

Deterioration in emotional response to depressive symptoms was defined as a pre-post change score of at least 3.1 points in the relating composite scale. The RCI was calculated using the standard deviation of this composite scale ($SD_{baseline}$ = 1.9), the pooled reliability coefficients from the Brief IPQ validation study (r_{tt} = 0.66), and a 95% confidence level.

Deterioration in suicidal ideation was defined as the pre-post change score of at least 1 point in the PHQ-9 suicide item.

Sample

We performed this secondary analysis in the per protocol sample which included 948 (81%) out of 1178 randomised participants who had at least one post-baseline value of one of the outcomes and no major protocol violation. Major protocol violations were pre-defined as not receiving or adhering to the intervention (i.e., feedback not opened, feedback reading time less than 15 seconds or no download of feedback form), multiple participation (post-hoc data check or self-report), reports of not having answered the survey seriously, baseline survey completion time less than two minutes and provision of an invalid email address. We preferred per protocol over intention-to-treat (ITT) sample, as the second is likely to underestimate the risk of an event by inflating the denominator with participants who have

provided invalid data or have never received the intervention. Whereas this is conservative in efficacy evaluations, in the current case of a risk evaluation we consider it more appropriate to prevent failing to detect a risk than overestimating it [43].

Additionally, we performed sensitivity analyses in the ITT sample, both with and without missing data imputation. We used two strategies for imputing data: assuming that all drop-outs were deteriorators, considering this to be the most conservative estimate (worst case); and assuming that all drop-outs were non-deteriorators, considering this to be the most optimistic estimate (best case).

Statistical analyses

We compared the rates of negative effects between study arms in terms of relative risks (RR). The RR estimates how much higher (or lower) the probability of negative effects is for participants in the respective feedback arm compared to the no feedback arm. To directly estimate the RR with 95% confidence intervals (CIs), we applied generalised linear models with a log link and robust sandwich variance estimator using modified log-Poisson regressions [44]. We chose this approach over alternative models as it is suited as well in case of frequent outcomes and suffers least from convergence problems [45, 46]. To test for differential effects in the subgroup of false positive screened participants, we ran another series of models additionally including false positive screens and the false positive screen by study arm interaction term. We set the significance level at α=0.05 and did not correct for multiple testing for two reasons: the trial was not powered for this secondary analysis, and as already mentioned, in case of negative effects we consider it more important to prevent the inflation of a type II error (i.e. failing to detect negative effects in case they exist) instead of the type I error. As some negative effects turned out to be rare in the study data, we also estimated odds ratios based on logistic regression models as post-hoc sensitivity analyses. We performed analyses with IBM SPSS version 27.

Results

Study flow and participant Characteristics

Of initially 5457 study participants, 4878 completed the screening questionnaire, and 1178 eligible participants were assigned to receive either no feedback (n = 391), non-tailored feedback (n = 393) or tailored feedback (n = 394) on their depression screening result. Of the 787 participants randomised to receive any feedback, in total 744 (95%) opened the feedback screen, of which 464 (62%) downloaded the PDF and 248 (33%) interacted with the feedback by clicking at least one link or modal. There was no descriptive difference between the

feedback engagement across feedback arms (see [12], for results per study arm). At one month, 976 participants provided follow-up data of outcome measures (loss to follow-up: 17%), of which 909 were included in the per protocol analysis. At six months, 965 participants provided follow-up data of outcome measures (loss to follow-up: 18%), of which 902 were included in the per protocol analysis. Numbers per study arm and analysis time point are shown in the CONSORT flow chart (Figure 2).

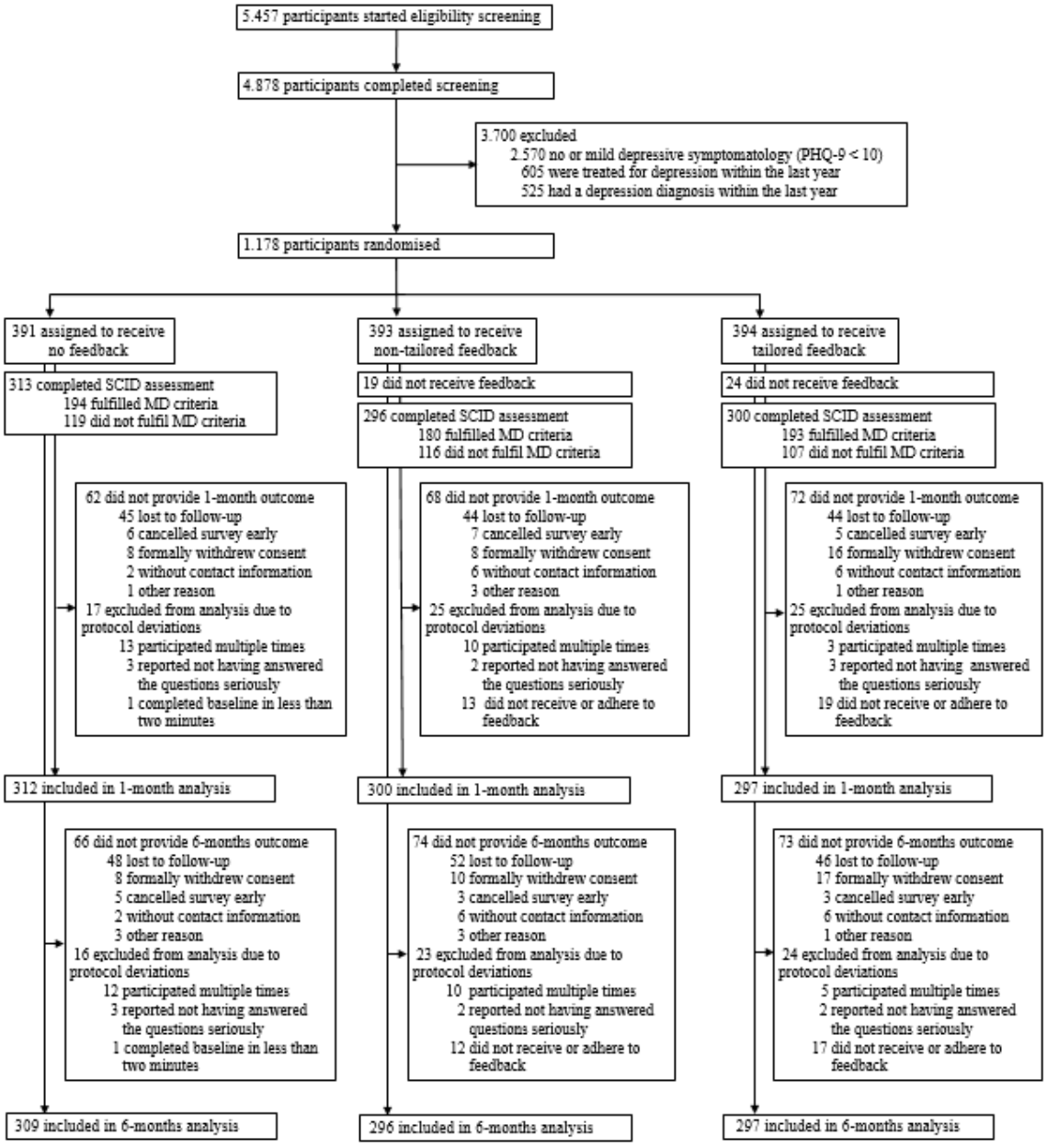


Figure 2. CONSORT flow chart (per protocol sample).

Relevant demographic and clinical characteristics of the per protocol sample were balanced across the three study arms (Table 1). The mean participant age was 37.3 (SD 14.1), 685 (72%) participants were women, and 488 (52%) had a high education level. At baseline, the average PHQ-9 depression severity score was 14.8 (SD 4.0), the average score in emotional response to depressive symptoms was 7.0 (SD 1.9), 455 participants (48%) reported to suffer from suicidal ideation at least several days within the last two weeks, and 820 (86%) thought that they currently suffered or maybe suffered from a depressive disorder. Out of 820 participants who were reached for diagnostic telephone interviews, 514 (63%) met the criteria for a major depression according to the DSM-5. Conversely, 306 participants (37%) were classified as false positive screens. Characteristics of the per protocol sample are comparable to those of the ITT sample (Multimedia Appendix 6).

Table 1. Baseline demographic and clinical characteristics of the per protocol sample.

	Total sample (*n* = 948)	Non-tailored feedback (*n* = 314)	Tailored feedback (*n* = 307)	No feedback (*n* = 327)
Age, years	37.3 (14.1)	37.8 (14.0)	36.8 (14.3)	37.2 (14.0)
Gender				
Female	685 (72%)	223 (71%)	219 (71%)	243 (74%)
Male	255 (27%)	88 (28%)	85 (28%)	82 (25%)
Divers	8 (0.8%)	3 (1.0%)	3 (1.0%)	2 (0.6%)
German mother tongue	902 (95%)	295 (94%)	296 (96%)	311 (95%)
Migration background	103 (11%)	30 (10%)	32 (10%)	41 (13%)
Being in a relationship	445 (47%)	162 (52%)	143 (47%)	140 (43%)
Living with others	631 (67%)	217 (69%)	202 (66%)	212 (65%)
Formal school education				
Low (less than 10 years)	160 (17%)	60 (19%)	44 (15%)	56 (17%)
Middle (at least 10 years)	300 (32%)	100 (32%)	102 (33%)	98 (30%)
High (A-level or above)	488 (52%)	154 (49%)	161 (52%)	173 (53%)
Working	691 (73%)	230 (73%)	232 (76%)	229 (70%)
Depression severity (PHQ-9)	14.8 (4.0)	14.9 (4.2)	14.6 (3.8)	14.8 (4.0)
Emotional response to depressive symptoms (composite scale)	7.0 (1.9)	7.0 (1.9)	6.9 (1.7)	6.9 (2.0)
Quality of life (EQ-5D-5L VAS)	57.6 (21.6)	57.2 (21.6)	58.2 (21.3)	57.4 (21.9)
Anxiety severity (GAD-7)	12.1 (4.3)	12.5 (4.2)	11.8 (4.3)	12.0 (4.3)
Somatic symptom severity (SSS-8)	14.4 (5.2)	14.5 (5.1)	14.2 (5.2)	14.4 (5.2)
No. of depression risk factors[a]	6.0 (2.4)	6.1 (2.5)	5.9 (2.3)	6.1 (2.5)
Frequency of suicidal ideation within last two weeks (PHQ-9 item 9)				
None	493 (52%)	161 (51%)	165 (54%)	167 (51%)
Several days	305 (32%)	113 (36%)	94 (31%)	98 (30%)
More than half the days	86 (9%)	23 (7%)	26 (9%)	37 (11%)
Nearly every day	64 (7%)	17 (5%)	22 (7%)	25 (8%)
Self-identifying as suffering from depression				
No	128 (14%)	32 (10%)	51 (17%)	45 (14%)
Maybe	432 (46%)	160 (51%)	141 (46%)	131 (40%)
Yes	388 (41%)	122 (39%)	115 (38%)	151 (46%)
Meeting criteria for major depression (SCID)	514 (63%)[b]	161 (61%)[c]	172 (64%)[d]	181 (62%)[e]

Data are mean (SD) or n (%). PHQ-9=Patient Health Questionnaire-9 (0 to 27). IPQ= Illness Perception Questionnaire (0 to 10). EQ-5D-5L=EuroQoL-5 Dimensions-5 Level scale (0 to 100). VAS=visual analogue scale. GAD-7=Generalized Anxiety Disorder-7 (0 to 21). SSS-8=Somatic Symptom Scale (0 to 32). SCID=Structured Clinical Interview for DSM-5 Disorders. [a]Risk factors included self-reported anxiety, addiction, traumatic life events, persistent physical symptoms, mood swings, chronic physical condition, lack of social support, mental comorbidity, mental comorbidity in family, history of suicide, current pregnancy, post-natal phase, menopause, premenstrual syndrome. [b]128 cases with missing data. [c]37 cases with missing data. [d]51 cases with missing data. [e]40 cases with missing data.

Negative effects outcomes

Misdiagnosis rates six months after randomisation were not higher after non-tailored (RR=1.30, 95% CI 0.59 to 2.86; P = .51) nor tailored feedback (RR=1.09, 95% CI 0.48 to 2.46; P = 0.84) as compared to no feedback, with rates of 4.9%, 4.1% and 3.5% in the non-tailored, the tailored, and the no feedback arm, respectively. Mistreatment rates six months after randomisation were not higher after non-tailored (RR=0.87, 95% CI 0.49 to 1.56; P = 0.65) nor tailored feedback (RR=0.95, 95% CI 0.54 to 1.67; P = 0.86), either, with rates of 7.2%, 7.7%, and 8.3% in the non-tailored, the tailored, and the no feedback arm. Rates of deterioration in depression severity were not higher after non-tailored (one month: RR=1.96, 95% CI 0.89 to 4.34; P = 0.1; six months: RR=0.60, 95% CI 0.3 to 1.19; P = 0.14) nor tailored feedback (one month: RR=0.70, 95% CI 0.25 to 1.94; P = 0.49; six months: RR=0.74, 95% CI 0.39 to 1.41; P = 0.37), with rates of 5.7%, 2.0%, and 2.9% at one month and 4.1%, 5.1%, and 6.8% at six months in the non-tailored, tailored, and no feedback study arm. Rates of deterioration in emotional response to depressive symptoms were not higher after non-tailored (one month: RR=1.18, 95% CI 0.43 to 3.21; P = 0.75; six months: RR=0.46, 95% CI 0.14 to 1.49; P = 0.2) nor tailored feedback (one month: RR=0.23, 95% CI 0.06 to 1.42; P = 0.13; six months: RR=0.70, 95% CI 0.25 to 1.94; P = 0.49) either, with rates of 2.7%, 0.7%, and 2.3% at one month and 1.4%, 2.0%, and 2.9% at six months. Rates of deterioration in suicidal ideation were not higher after non-tailored (RR=1.12, 95% CI 0.69 to 1.8; P = 0.66) or tailored feedback (RR=1.40, 95% CI 0.39 to 1.41; P = 0.15) at six months, with rates of 10.5%, 13.1%, and 9.4%. At one month, however, the rate of deterioration in suicidal ideation was almost two-fold higher in the non-tailored (RR=1.92; 95% CI 1.14 to 3.24; P = 0.01), but not in the tailored feedback arm (RR=1.26, 95% CI 0.25 to 1.94; P = 0.43), as compared to no feedback. Rates in the non-tailored, the tailored, and the no feedback arm were 12.3%, 8.1%, and 6.4%. Absolute frequencies and rates for all negative effects per study arm and time point are shown in Table 2. Relative risks with corresponding 95% CIs are illustrated in Figure 3.

Table 2. Absolute frequencies and rates of negative effects per study arm and time point.

	n	Non-tailored feedback	n	Tailored feedback	n	No feedback
Misdiagnosis (6 months) [a]	263	13 (4.9%)	267	11 (4.1%)	290	11 (3.5%)
Mistreatment (6 months) [a]	263	19 (7.2%)	267	21 (7.7%)	290	24 (8.3%)
Psychotherapy [a]	263	13 (4.9%)	267	17 (6.4%)	290	18 (6.2%)
Medication [a]	263	9 (3.4%)	267	6 (2.2%)	290	8 (2.8%)
Deterioration in depression severity						
1 month	300	17 (5.7%)	297	6 (2.0%)	312	9 (2.9%)
6 months	296	12 (4.1%)	297	15 (5.1%)	309	21 (6.8%)
Deterioration in suicidal ideation						
1 month	300	37 (12%)	297	24 (8.1%)	312	20 (6.4 %)
6 months	296	31 (11%)	297	39 (13%)	309	29 (9.4%)
Deterioration in emotional response						
1 month	299	8 (2.7%)	296	2 (0.7%)	308	7 (2.3%)
6 months	294	4 (1.4%)	299	6 (2.0%)	307	9 (2.9%)

Data are n (%). [a]n refers to participants who completed both the follow-up assessment and the SCID depression module at baseline.

Results did not differ for the subgroup of false-positives ($P_{interaction}$ ranging between 0.29 and 0.8). Sensitivity analyses based on logistic regression models as well as those in the ITT sample with the full analysis set and with missing data imputation based on the best case scenario showed comparable results. In the ITT analysis based on the worst case scenario, however, the relative risk for deterioration in suicidal ideation in the non-tailored feedback arm at one month was not higher than in the no feedback arm (RR=1.26, 95% CI 0.99 to 1.61; P = 0.07; Multimedia Appendix 7). Based on exploratory post hoc analyses, baseline demographic and clinical characteristics of all participants deteriorated in any outcome at any time point were comparable to the total sample. (Multimedia Appendix 8).

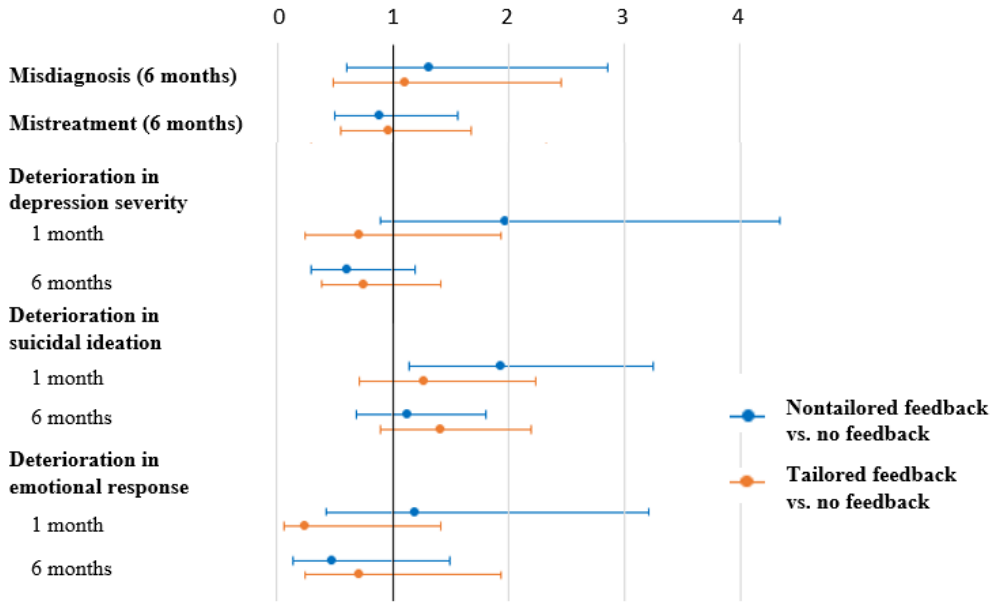


Figure 3. Relative risks (95% CIs) for all negative effects at all time points in the non-tailored and tailored feedback arm as compared to no feedback.

Discussion

To the best of our knowledge, this secondary analysis is the first study to systematically examine potential negative effects of feedback after internet-based depression screening in a large sample of currently undiagnosed and untreated individuals with at least moderate depression severity.

Summary of results

The results indicate that feedback, both non-tailored and tailored, was not associated with increased rates of misdiagnosis, potential mistreatment, deterioration in depressive symptoms, or deterioration in emotional response to symptoms as compared to no feedback. Deterioration of suicidal ideation, however, appeared to be more likely one month after receiving non-tailored feedback compared with no feedback. Although almost 40% of the sample turned out to be screened false positive, irrespective of the study arm rates of subsequent misdiagnosis and potential mistreatment were lower than 5% and 9%, respectively, with rates of pharmacotherapy ranging even lower than 4%. Across study arms, deterioration in emotional response to depressive symptoms was reported by at most 3% of participants, deterioration of depression severity by at most 7%, and deterioration of suicidal ideation by at most 13%.

Limitations

The interpretation of these results should be considered in the context of the study's limitations. First, the underlying DISCOVER trial did not explicitly call for those seeking depression screening. As these may be more eager to follow the advice of the feedback, in this sample misdiagnosis and potential mistreatment might be underestimated. Second, the selection of outcomes was limited and relevant negative effects such distress, stigma, suicidal behaviours, treatment side effects, or overdiagnosis (i.e. the diagnosis of correctly diagnosed but mild cases that would not benefit from treatment [15]) could not be assessed. Third, assessments of outcoms were based on self-reports and have limitations. Regarding mistreatment, it cannot be ruled out that participants (correctly) received antidepressant medication or psychotherapy for conditions other than depressive disorders, wherefore rates may be overestimated. Further, the operationalisations of suicidal ideation and emotional response to depressive symptoms are based on a single item and a composite score not well validated for this purpose (see e.g., [47, 48] for research on the validity of the PHQ-9 suicide item). Lastly, this secondary analysis of the DISCOVER trial was planned post-hoc and therefore not powered to detect differences between groups regarding selected outcomes, and multiple testing might have led to overestimation of significance in the case of deterioration in suicidal ideation. Notably, the findings refer to the German health care system where psychotherapy is available and covered by the social health insurance. Particularly rates for misdiagnosis and mistreatment might differ in other countries with differing health policies.

Principal findings, comparison with prior work, and implications

Taken together, there are three main findings. First, the results regarding mistreatment and misdiagnosis emphasise that feedback after internet-based depression screening is not associated with inadequate management and care for individuals who receive false positive feedback - even when the rate of false positives is relatively high and when the feedback refers to a health system that covers depression care. Extending on prior findings that feedback after internet-based depression screening does not affect service uptake (KOHLMANN, [49]), these results refute one postulated but opinion-based criticism against internet-based depression screening [4, 14]. Second, there is also no indication that feedback after internet-based depression screening induces negative psychological effects such as deterioration in depression severity and emotional response to symptoms. Notably, the rates for deterioration in depression severity of at most 7% found in this study are comparable to those reported in care as usual conditions in psychotherapy trials [50]. The null findings regarding deterioration in emotional reponse to symptoms, however, appear to conflict with

prior qualitative evidence suggesting that internet-based depression screening does induce negative emotions and distress in some individuals [8, 17]. An explanation for this discrepancy might be that negative emotional effects might be induced not only by the feedback but also by the screening questions alone, which has been reported in a qualitative follow-up study of the DISCOVER trial [17]. Furthermore,it might be that the construct emotional response, defined by items assessing concern and emotional affectedness about the symptoms, relates more to a cognitive evaluation of symptoms rather than capturing an actual emotional state. Therefore, comparing outcomes such as distress or negative affectivity shortly after providing the screening vs. the feedback appears worthwile to further address these issues (see [51, 52] for examplary study designs in suicide screening). Third, the current results do not rule out that non-tailored feedback, in contrast to tailored feedback, might lead to increased suicidal ideation after one month. This finding is contradictory to results from a randomised clinical trial on screening and feedback in the primary care setting [53], but in line with prior observational evidence regarding internet-based screeners [18]. Explanations for such an effect might be that receiving a diagnosis online might induce hopelessness, a known risk factor for suicidal ideation [54], or that the referral initiation process may be overwhelming, thereby triggering decompensation [18]. However, it remains an open question why non-tailored but not tailored feedback should increase suicidal ideation: against our hypothesis, neither the usage of the feedback nor any other outcome differed between the two feedback arms [12]. Further, increased suicidal ideation was not reported by participants when qualitatively asked for adverse events six months after randomisation via telephone (see [12] for main results). Explanations for this discrepancy might be that reporting of suicidal ideation might have been stigmatising, might be not remembered or not classified as an adverse event by participants. Given that these results should be interpreted with caution due to the study's limitations, more robust research is needed to further address suicidal ideation in internet-based depression screening. If prospective trials that use validated ouctome measures corroborate an association of internet-based screening and/or feedback with suicidal ideation, this should inform regulations of currently unmonitored internet-based depression tests. Further, the findings should also inform research regarding comparable depression screening in medical and primary care settings, which is currently recommended in many countries despite very uncertain evidence regarding potential harms [55].

Conclusion

In conclusion, the results of this secondary analysis indicate that feedback after internet-based depression screening is neither associated with healthcare related negative effects such as

misdiagnosis and potential mistreatment nor with psychological negative effects such as deterioration in depression severity or emotional response to symptoms. However, it cannot be ruled out that non-tailored feedback may be associated with increased suicidal ideation. Against the background of the study's secondary design, robust prospective research on negative effects and particularly suicidal ideation in internet-based depression screening is needed to inform current practice of public internet-based depression screening as well as research in the field of depression screening in general.

Funding

This work was funded by the German Research Foundation as part of the underlying DISCOVER RCT (grant number: 424162019).

Contributions

FS, AD, and SK designed the study. SK and BL obtained the funding for the underlying DISCOVER trial. FS and SK collected the data of the DISCOVER trial. FS performed the data analyses and wrote the original draft of the manuscript. All authors contributed to the interpretation of the data for the paper and critically reviewed and edited the original draft. All authors had full access to all the data in the study, approved the final version of the manuscript, and take responsibility for its submission for publication.

Conflicts of Interest

FS declares that there is no conflict of interest.

SK reports research funding (no personal honoraria) from the German Research Foundation and the German Federal Ministry of Education and Research.

BL reports research funding (no personal honoraria) from the German Research Foundation, the German Federal Ministry of Education and Research, the German Innovation Committee at the Joint Federal Committee, the European Commission Horizon 2020 Framework Programme, the European Joint Programme for Rare Diseases (EJP), the Ministry of Science, Research and Equality of the Free and Hanseatic City of Hamburg, Germany and the Foundation Psychosomatics of Spinal Diseases, Stuttgart, Germany. He has received remuneration for several scientific book articles from various book publishers and as a committee member from Aarhus University, Denmark. He received travel expenses from the European Association of Psychosomatic Medicine (EAPM) and accommodation and meals from the Societatea de Medicina Biopsyhosociala, Romania, for a presentation at the EAPM Academy at the Conferința Națională de Psihosomatică, Cluj-Napoca, Romania, Oct 2023. He

was a board member of the European Association of Psychosomatic Medicine (EAPM) (unpaid) until 2022.

AD reports research funding (no personal honoraria) from the German Research Foundation, the German Federal Ministry of Education and Research, the German Innovation Committee at the Joint Federal Committee, and German Cancer Aid.

Acknowledgements

We would like to thank Professor Dr Levente Kriston for his critical and constructive comments regarding data analysis.

Abbreviations

CI: confidence interval
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
EQ-5D-5L: EuroQoL-5 Dimensions-5 Level scale
GAD-7: Generalized Anxiety Disorder-7
IPQ: Illness Perception Questionnaire
ITT: intention-to-treat
PHQ-9: Patient Health Questionnaire-9
RCI: reliable change index
RR: relative risk
SCID: Structured Clinical Interview for DSM-5 Disorders
SSS-8: Somatic Symptom Scale-8
VAS: visual analogue scale

Multimedia Appendix 1

CONSORT-EHEALTH V1.6.

Multimedia Appendix 2

Illustration of the digitised PHQ-9

Multimedia Appendix 3

Illustrations of the complete non-tailored feedback

Multimedia Appendix 4

Illustrations of the complete tailored feedback

Multimedia Appendix 5

Illustration of suicidal ideation feedback

Multimedia Appendix 6

Characteristics of ITT sample

Multimedia Appendix 7

Sensitivity analyses

Multimedia Appendix 8

Post hoc analyses

References

1. Collaborators, G.B.D.M.D., *Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019.* Lancet Psychiatry, 2022. **9**(2): p. 137-150.
2. Ghio, L., et al., *Duration of untreated illness and outcomes in unipolar depression: a systematic review and meta-analysis.* J Affect Disord, 2014. **152-154**: p. 45-51.
3. Barry, M.J., et al., *Screening for Depression and Suicide Risk in Adults: US Preventive Services Task Force Recommendation Statement.* Jama, 2023. **329**(23): p. 2057-2067.
4. Duckworth, K. and S. Gilbody, *Should Google offer an online screening test for depression?* BMJ, 2017. **358**: p. j4144.
5. Vaidyanathan, U., et al., *An evaluation of Internet searches as a marker of trends in population mental health in the US.* Sci Rep, 2022. **12**(1): p. 8946.
6. Eichenberg, C., C. Wolters, and E. Brähler, *The Internet as a Mental Health Advisor in Germany-Results of a National Survey.* Plos One, 2013. **8**(11): p. e79206.
7. Jacobson, N.C., et al., *Impact of online mental health screening tools on help-seeking, care receipt, and suicidal ideation and suicidal intent: Evidence from internet search behavior in a large U.S. cohort.* J Psychiatr Res, 2022. **145**: p. 276-283.
8. Kruzan, K.P., et al., *"I Wanted to See How Bad it Was": Online Self-screening as a Critical Transition Point Among Young Adults with Common Mental Health Conditions.* Proc SIGCHI Conf Hum Factor Comput Syst, 2022. **2022**.
9. Daniela, R.J., et al., *CONSORT Harms 2022 statement, explanation, and elaboration: updated guideline for the reporting of harms in randomised trials.* BMJ, 2023. **381**: p. e073725.
10. Rozental, A., et al., *Consensus statement on defining and measuring negative effects of Internet interventions.* Internet Interventions, 2014. **1**(1): p. 12-19.

11. Batterham, P.J., et al., *Online screening and feedback to increase help-seeking for mental health problems: population-based randomised controlled trial.* British Journal of Psychiatry Open, 2016. **2**(1): p. 67-73.
12. Kohlmann S, Sikorski F, König HH, Schütt M, Zapf A, Löwe B. *The efficacy of automated feedback after internet-based depression screening (DISCOVER): an observer-masked, three-armed, randomised controlled trial in Germany*. Lancet Digital Health. 2024. **6**(7): e446-e457.
13. Thombs, B.D., et al., *Screening for depression and anxiety in general practice.* BMJ, 2023. **382**: p. p1615.
14. Danczak, A., *Online screening test for depression is inappropriate.* BMJ, 2017. **359**: p. j4736.
15. Thombs, B., K.A. Turner, and I. Shrier, *Defining and Evaluating Overdiagnosis in Mental Health: A Meta-Research Review.* Psychother Psychosom, 2019. **88**(4): p. 193-202.
16. Ryan, A. and S. Wilson, *Internet healthcare: do self-diagnosis sites do more harm than good?* Expert Opinion on Drug Safety, 2008. **7**(3): p. 227-229.
17. Sikorski, F., et al., *The efficacy of automated feedback after internet-based depression screening: Study protocol of the German, three-armed, randomised controlled trial DISCOVER.* Internet Interventions, 2021. **25**: p. 100435.
18. Jacobson, N.C., et al., *Impact of online mental health screening tools on help-seeking, care receipt, and suicidal ideation and suicidal intent: Evidence from internet search behavior in a large U.S. cohort.* Journal of Psychiatric Research, 2022. **145**: p. 276-283.
19. The Efficacy of Automated Feedback After Internet-based Depression Screening (DISCOVER). ClinicalTrials.gov identifier: NCT04633096. Updated September 21, 2023. Accessed October 21, 2023. https://clinicaltrials.gov/study/NCT04633096
20. Center for Open Science. Does internet-based depression screening with feedback of results cause harm? A secondary analysis using data from the randomised controlled DISCOVER trial. May 2, 2023. Accessed October 21, 2023. https://osf.io/tzyrd
21. Boutron, I., et al., *CONSORT Statement for Randomized Trials of Nonpharmacologic Treatments: A 2017 Update and a CONSORT Extension for Nonpharmacologic Trial Abstracts.* Ann Intern Med, 2017. **167**(1): p. 40-47.
22. Eysenbach, G., *CONSORT-EHEALTH: implementation of a checklist for authors and editors to improve reporting of web-based and mobile randomized controlled trials.* Studies in Health Technology and Informatics, 2013. **192**: p. 657-61.
23. Moher, D., et al., *CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials.* BMJ, 2010. **340**: p. c869.
24. Montgomery, P., et al., *Reporting randomised trials of social and psychological interventions: the CONSORT-SPI 2018 Extension.* Trials, 2018. **19**(1): p. 407.
25. Schulz, K.F., D.G. Altman, and D. Moher, *CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials.* BMC Medicine, 2010. **8**(1): p. 18.
26. Kroenke, K., R.L. Spitzer, and J.B.W. Williams, *The PHQ-9.* Journal of General Internal Medicine, 2001. **16**(9): p. 606-613.
27. *DISCOVER website.*
28. Rohde, F., et al., *Optimization of the Mainzelliste software for fast privacy-preserving record linkage.* Journal of Translational Medicine, 2021. **19**(1): p. 1-12.
29. Löwe, B., et al., *Gesundheitsfragebogen für Patienten (PHQ-D). Manual und Testunterlagen*. 2002, Pfizer New York, NY.

30. Negeri, Z.F., et al., *Accuracy of the Patient Health Questionnaire-9 for screening to detect major depression: updated systematic review and individual participant data meta-analysis.* Bmj, 2021. **375**: p. n2183.
31. Du, N., et al., *Validity study of Patient Health Questionnaire-9 items for Internet screening in depression among Chinese university students.* Asia Pac Psychiatry, 2017. **9**(3).
32. Erbe, D., et al., *Interformat reliability of the patient health questionnaire: Validation of the computerized version of the PHQ-9.* Internet Interv, 2016. **5**: p. 1-4.
33. BÄK, KBV, and AWMF, *Nationale VersorgungsLeitlinie Unipolare Depression – Langfassung, Version 3.0.* . 2022.
34. Seeralan, T., et al., *Patient involvement in developing a patient-targeted feedback intervention after depression screening in primary care within the randomized controlled trial GET.FEEDBACK.GP.* Health Expect, 2020.
35. Kohlmann, S., et al., *Depression screening using patient-targeted feedback in general practices: study protocol of the German multicentre GET.FEEDBACK.GP randomised controlled trial.* BMJ Open, 2020. **10**(9): p. e035973.
36. Beesdo-Baum, K., M. Zaudig, and H.-U. Wittchen, *SCID-5-CV Strukturiertes Klinisches Interview für DSM-5-Störungen–Klinische Version: Deutsche Bearbeitung des Structured Clinical Interview for DSM-5 Disorders–Clinician Version von Michael B. First, Janet BW Williams, Rhonda S. Karg, Robert L. Spitzer*. 2019: Hogrefe.
37. Löwe, B., et al., *Monitoring depression treatment outcomes with the patient health questionnaire-9.* Medical Care, 2004. **42**(12): p. 1194-1201.
38. Broadbent, E., et al., *The brief illness perception questionnaire.* Journal of Psychosomatic Research, 2006. **60**(6): p. 631-637.
39. Jacobson, N.S. and P. Truax, *Clinical significance: A statistical approach to defining meaningful change in psychotherapy research.* Journal of Consulting and Clinical Psychology, 1991. **59**: p. 12-19.
40. Kroenke, K., R.L. Spitzer, and J.B. Williams, *The PHQ-9: validity of a brief depression severity measure.* J Gen Intern Med, 2001. **16**(9): p. 606-13.
41. Löwe, B., et al., *Monitoring depression treatment outcomes with the patient health questionnaire-9.* Med Care, 2004. **42**(12): p. 1194-201.
42. McMillan, D., S. Gilbody, and D. Richards, *Defining successful treatment outcome in depression using the PHQ-9: a comparison of methods.* J Affect Disord, 2010. **127**(1-3): p. 122-9.
43. Phillips, R., et al., *Analysis and reporting of adverse events in randomised controlled trials: a review.* BMJ Open, 2019. **9**(2): p. e024537.
44. Zou, G., *A Modified Poisson Regression Approach to Prospective Studies with Binary Data.* American Journal of Epidemiology, 2004. **159**(7): p. 702-706.
45. Gallis, J.A. and E.L. Turner, *Relative Measures of Association for Binary Outcomes: Challenges and Recommendations for the Global Health Researcher.* Ann Glob Health, 2019. **85**(1): p. 137.
46. Knol, M.J., et al., *Overestimation of risk ratios by odds ratios in trials and cohort studies: alternatives to logistic regression.* Cmaj, 2012. **184**(8): p. 895-9.
47. Na, P.J., et al., *The PHQ-9 Item 9 based screening for suicide risk: a validation study of the Patient Health Questionnaire (PHQ)−9 Item 9 with the Columbia Suicide Severity Rating Scale (C-SSRS).* Journal of Affective Disorders, 2018. **232**: p. 34-40.
48. Rossom, R.C., et al., *Suicidal ideation reported on the PHQ9 and risk of suicidal behavior across age groups.* Journal of Affective Disorders, 2017. **215**: p. 77-84.

49. Batterham, P.J., et al., *Online screening and feedback to increase help-seeking for mental health problems: population-based randomised controlled trial.* BJPsych Open, 2016. **2**(1): p. 67-73.
50. Cuijpers, P., et al., *The effects of psychotherapies for depression on response, remission, reliable change, and deterioration: A meta-analysis.* Acta Psychiatr Scand, 2021. **144**(3): p. 288-299.
51. Gould, M.S., et al., *Evaluating iatrogenic risk of youth suicide screening programs: a randomized controlled trial.* JAMA, 2005. **293**(13): p. 1635-43.
52. Robinson, J., et al., *Does screening high school students for psychological distress, deliberate self-harm, or suicidal ideation cause distress--and is it accepTable? An Australian-based study.* Crisis, 2011. **32**(5): p. 254-63.
53. Löwe, B., et al., *Clinical effectiveness of patient-targeted feedback following depression screening in general practice (GET.FEEDBACK.GP): an investigator-initiated, prospective, multicentre, three-arm, observer-blinded, randomised controlled trial in Germany.* The Lancet Psychiatry, 2024. **11**(4): p. 262-273.
54. Kuo, W.-H., J.J. Gallo, and W.W. Eaton, *Hopelessness, depression, substance disorder, and suicidality.* Social Psychiatry and Psychiatric Epidemiology, 2004. **39**(6): p. 497-501.
55. Beck, A., et al., *Screening for depression among the general adult population and in women during pregnancy or the first-year postpartum: two systematic reviews to inform a guideline of the Canadian Task Force on Preventive Health Care.* Systematic Reviews, 2022. **11**(1): p. 176.

7.3.1. Publication III: Supplementary Material

Multimedia Appendix 1

CONSORT-EHEALTH V1.6. (p.129)

Multimedia Appendix 2

Illustration of the digitised PHQ-9 (p.141)

Multimedia Appendix 3

Illustrations of the complete non-tailored feedback; see section 7.2.1. Publication II: Supplementary Material; supplementary Figure 1 (p. 75).

Multimedia Appendix 4

Illustrations of the complete tailored feedback; see section 7.2.1. Publication II: Supplementary Material; supplementary Figure 2 (pp. 76-79).

Multimedia Appendix 5

Illustration of suicidal ideation feedback; see section 7.2.1. Publication II: Supplementary Material; supplementary Figure 3 (p. 80).

Multimedia Appendix 6

Characteristics of ITT sample (p. 142)

Multimedia Appendix 7

Sensitivity analyses (p. 144)

Multimedia Appendix 8

Post hoc analyses (p. 147)

Multimedia Appendix: CONSORT-EHEALTH Checklist V1.6.2 Report

(based on CONSORT-EHEALTH V1.6.1 from: CONSORT-EHEALTH (V 1.6.1) - Submission/Publication Form (google.com); Eysenbach G, CONSORT-EHEALTH GroupCONSORT-EHEALTH: Improving and Standardizing Evaluation Reports of Web-based and Mobile Health InterventionsJ Med Internet Res 2011;13(4):e126

Date completed // updated

17/2/2024 // 12/04/2024

by

Franziska Sikorski

Does feedback after internet-based depression screening cause harm? A secondary analysis of negative effects in the randomised controlled DISCOVER trial

Language

German

Accessibility; URL of your Intervention Website or App

exemplary access is possible (feedback without screening); https://www.discover-studie.de/rueckmeldung

Primary Medical Indication/Disease/Condition

undiagnosed depressive disorders

Primary Medical Indication/Disease/Condition

mistreatment, misdiagnosis, deterioration in depression severity, deterioration in emotional response to symptoms, deterioration in suicidal ideation

Overall, was the intervention effective?

potentially harmful: increased deterioration in suicidal ideation in tailored feedback arm

TITLE

1a) Does your paper address CONSORT item 1a (identification as a randomized trial in the title)?

yes

1a-i) Identify the mode of delivery in the title

“Does feedback after web-based depression screening cause harm? A secondary analysis of negative effects in the randomised controlled DISCOVER trial”

1a-ii) Non-web-based components or important co-interventions in title

This is not relevant to this manuscript.

1a-iii) Primary condition or target group in the title

The intervention is not addressing a specific condition (general population; with undiagnosed depressive symptoms).

ABSTRACT

1b-i) Key features/functionalities/components of the intervention and comparator in the METHODS section of the ABSTRACT

"Undiagnosed but affected individuals with a positive depression screening score (Patient Health Questionnaire-9, PHQ-9 ≥ 10 points) were randomised to receive no feedback, nontailored feedback, or tailored feedback on their screening result together with recommendations to seek diagnostic advice."

1b-ii) Level of human involvement in the METHODS section of the ABSTRACT

"We aimed to examine whether automated feedback after internet-based depression screening…"

1b-iii) Open vs. closed, web-based (self-assessment) vs. face-to-face assessments in the METHODS section of the ABSTRACT

"Participants were followed-up at one and six months via online questionnaires. Misdiagnosis and mistreatment were operationalised as having received a depression diagnosis by a health professional and as having started psychotherapy or antidepressant medication since screening while not meeting the DSM-5 criteria of a major depression (diagnostic telephone interviews). Deterioration in depressive symptoms was defined as a pre-post change of ≥ 4.4 points in the PHQ-9, deterioration in emotional response to symptoms as a pre-post change of ≥ 3.1 points in a composite scale based on the Brief Illness Perception Questionnaire, and deterioration in suicidal ideation as a pre-post change of ≥ 1 point in the PHQ-9 suicide item."

1b-iv) RESULTS section in abstract must contain use data

"In the per protocol sample of 948 participants who opened the feedback…"

1b-v) CONCLUSIONS/DISCUSSION in abstract for negative trials

Not applicable, as all participants used the intervention.

INTRODUCTION

2a-i) Problem and the type of system/solution

"Depressive disorders, although being among the most disabling and most prevalent disorders worldwide [1], often remain undetected and therefore untreated [2]. In the last decades, depression screening has been increasingly discussed as promising to reach those affected but undetected at an early stage: In addition to population level screening in routine clinical care, as for example recommended in the United States [3], advocates also speak out in favour of screening for depression online [4]. For many affected individuals, the internet is already the favoured source for information on mental health [5, 6]. Further, so called internet-based depression tests are widely promoted by mental health-related institutions and frequently used by those seeking diagnostic advice [7]. The rational of internet-based depression tests typically involves administering symptom-based screening questionnaires and providing individuals with direct feedback on screening results, sometimes supplemented by links or referrals to services. The feedback is thought to empower affected individuals to better act on their symptoms [8] and to seek diagnostic consultation and, if necessary, appropriate care. As such, it might improve early detection and management of depression."

2a-ii) Scientific background, rationale: What is known about the (type of) system

“Negative effects, if present, would therefore likely be genereated without creating substantial health benefits. Evidence in this regard is, however, scarce, with the current scientific debate being mainly reflected by opinion papers: The first area of negative effects of depression screening, discussed in both medical and internet-based contexts, relates to inadequate management and care for individuals who receive false positive feedback. Critics particularly point to the risk of increased rates of misdiagnosis and mistreatment following screening. This, again, is assumed to lead to unnecessary iatrogenic effects such as adverse medication and psychotherapy side effects in healthy individuals, societal costs, and waste of limited health care resources resulting in potential undertreatment of more severe cases [4, 12, 13]. A second area of concern relates to negative psychological effects to the feedback of screening results. It is assumed that the associated labeling, resembling a clinical diagnosis, might induce anxiety, distress, stigma, or nocebo effects such as for example deterioration of symptoms [4, 13, 14]. These effects could be amplified by the fact that, in contrast to medical settings, in internet-based depression screening the ‘diagnosis’ would be delivered without a health professional who could provide emotional support or advice on further steps [15]. Indeed, in qualitative studies on internet-based mental health screening some participants describe having been discouraged, shocked or concerned by the feedback they received [8, 16]. Further, one observational study found that screening procedures including

referrals to in-person care had a higher likelihood of subsequent online searches for suicidal intent, potentially suggesting a deterioration of suicidal ideation [17]."

Does your paper address CONSORT subitem 2b?

"In the present study, we addressed this lack of evidence by analysing data from our recently conducted randomised controlled trial on the efficacy of feedback after internet-based depression screening (cite paper, when published). Based on the potential negative effects discussed in the literature and on outcomes assessed in that trial, we aimed to examine whether feedback after internet-based depression screening is associated with increased misdiagnosis, mistreatment, deterioration in depressive symptoms, deterioration in emotional response to symptoms, and deterioration in suicidal ideation one and six months after the screening."

METHODS

3a) CONSORT: Description of trial design (such as parallel, factorial) including allocation ratio

"DISCOVER was an investigator-initiated, observer-blinded, three-armed, randomised controlled trial that compared automated feedback with no feedback after internet-based depression screening. After being screened for depression with the digitised Patient Health Questionnaire-9 (PHQ-9 [25]), eligible participants were randomised to receive either no feedback, nontailored feedback or tailored feedback on their screening result (1:1:1 allocation ratio)."

3b) CONSORT: Important changes to methods after trial commencement (such as eligibility criteria), with reasons

"We conducted small deviations from the preregistration: we added the outcomes misdiagnosis and emotional response to symptoms, as we deemed this of clinical interest. Further, we added sensitivity analyses based on logistic regression models."

3b-i) Bug fixes, Downtimes, Content Changes

We did not have major bugs or down time for this trial.

4a) CONSORT: Eligibility criteria for participants

"Participants were individuals aged 18 years or above with at least moderate depression severity (PHQ-9 ≥ 10) but not diagnosed with or treated for depression within the last year. Additional eligibility criteria were having sufficient internet literacy and German language proficiency, providing contact details, and giving online informed consent."

4a-i) Computer / Internet literacy

Reported, see 4a).

4a-ii) Open vs. closed, web-based vs. face-to-face assessments:

"The study was promoted as being on 'stress and psychological well-being' on a publicly accessible study website [26]. The aim of evaluating internet-based depression screening was not explicitly communicated, but interested participants were informed that some of them will get feedback on a part of their answers. Traditional and social media campaigns as well as print advertisements in public areas of several German cities were used to approach interested individuals. To reach a sample that strives for representativeness of the German population with respect to age and gender, a marketing company further advertised the study via a nationwide internet-based access survey panel. "

"Double entries identified based on personal data by a privacy-preserving record linkage service [27] were automatically re-allocated to their former study arm. Research staff were masked to allocation at any time until breaking the blind."

"Web-based follow-up assessments were set at one month and six months after randomisation. Two to five days and six months after randomisation, participants were contacted via telephone for complementary diagnostic interviews."

4a-iii) Information giving during recruitment

"The study was promoted as being on 'stress and psychological well-being' on a publicly accessible study website [26]. The aim of evaluating internet-based depression screening was not explicitly communicated, but interested participants were informed that some of them will get feedback on a part of their answers."

4b) CONSORT: Settings and locations where the data were collected

"The study was promoted as being on 'stress and psychological well-being' on a publicly accessible study website [26]."

"Traditional and social media campaigns as well as print advertisements in public areas of several German cities were used to approach interested individuals across Germany."

4b-i) Report if outcomes were (self-)assessed through online questionnaires

"Web-based follow-up assessments were set at one month and six months after randomisation, with up to ten automatic email reminders being sent to participants in case of incomplete surveys. "

4b-ii) Report how institutional affiliations are displayed

Figure 1 shows how the logos of the University Medical Center Hamburg and of the German Research Foundation were displayed to participants.

5) CONSORT: Describe the interventions for each group with sufficient details to allow replication, including how and when they were actually administered

5-i) Mention names, credential, affiliations of the developers, sponsors, and owners

The software was developed by the authors together with an IT specialist. The software is not commercially available.

5-ii) Describe the history/development process

"The feedback was developed in a multistage process involving patient representatives [33, 34], an IT specialist, and a digital graphic agency to adapt the material to the possibilities of internet-based presentation."

5-iii) Revisions and updating

The intervention was not revised during the trial and only this original version was deployed.

5-iv) Quality assurance methods

"Double entries identified based on personal data by a privacy-preserving record linkage service [27] were automatically re-allocated to their former study arm."

5-v) Ensure replicability by publishing the source code, and/or providing screenshots/screen-capture video, and/or providing flowcharts of the

"Participants of the feedback arms received information on follow-up procedures and were offered feedback on their screening result by clicking on a 'next'-button (Figure 1)."

"Illustrations of the complete nontailored and tailored feedback versions can be found in supplements B and C."

5-vi) Digital preservation

The feedback is accessible (www.discover-studie.de/rueckmeldung; www.discover-studie.de/personalisierte-rueckmeldung); the intervention is however not archived.

5-vii) Access

"The study was promoted as being on 'stress and psychological well-being' on a publicly accessible study website [26]. "

5-viii) Mode of delivery, features/functionalities/components of the intervention and comparator, and the theoretical framework

"After completing the baseline survey, all participants were thanked for participating in the study. Participants of the feedback arms received information on follow-up procedures and were offered feedback on their screening result by clicking on a 'next'-button (Figure 1). Both nontailored and tailored feedback comprised four sections: (1) the depression screening result, (2) a note to seek diagnostic consultation by a health professional together with a link to make an appointment within the next two weeks, (3) brief general information on depression, and (4) information on depression treatment based on the German National Clinical Practice Guideline for Unipolar Depression [32]. Notably, in the German health care system depression care is available and covered by the social health insurance. Information was extended by direct links to referenced health or social services (e.g. web-based therapies covered by the health insurance, self-help groups), and the feedback form could be downloaded in a file that included all active links. In extension to the nontailored feedback, the information in the tailored feedback intervention was personalised to participants' characteristics (e.g., 'You have indicated that you had low spirits, sleep disturbances, and loss of energy during the past two weeks.'). Additionally, after being provided with the screening result (section 1) but before receiving further information (sections 2 to 4), participants were asked whether they think that their symptoms were indications of depression and whether they worried about the symptoms. According to the participants' answers, the following three feedback sections were arranged in a differing order, phrased slightly differently, and extended by information tailored to participants' risk profile (e.g. 'Depression in pregnancy is common.'). The feedback was developed in a multistage process involving patient representatives [33, 34] and an IT specialist and a digital graphic agency to adapt the material to the possibilities of web-based presentation. Illustrations of the complete nontailored and tailored feedback versions can be found in Multimedia Appendices 3 and 4."

5-ix) Describe use parameters

This intervention is a one-time use intervention.

5-x) Clarify the level of human involvement

There was no human involvement in the feedback interventions. There was only huan involvement in the telephone assessments, as already mentioned.

5-xi) Report any prompts/reminders used

"Internet-based follow-up assessments were set at one month and six months after randomisation, with up to ten automatic email reminders being sent to participants in case of incomplete surveys. Two to five days and six months after randomisation, participants were contacted via telephone for complementary diagnostic interviews, with calls being repeated at different hours during daytime and evening in case participants were not reached (see [16] for more detailed information on the data collection). "

5-xii) Describe any co-interventions (incl. training/support)

There are no co-interventions in this trial.

6a) CONSORT: Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed

Measures

"Depression diagnosis by a health professional was assessed at six months with the question: "Have you been diagnosed with depression or burnout in the last six months?".Guideline-based depression treatment, i.e. psychotherapy and/or pharmacotherapy with antidepressant medication recommended by the German National Clinical Practice Guideline for Unipolar Depression [32], was assessed at six months with the questions: "Have you started any psychotherapy or similar treatment in the last 6 months [which]?") and "Have you started taking medication to treat depression or other complaints such as sleep problems, anxiety or stress [which ones]?". Participants could choose from guideline-based treatment options or give open answers. In case of open answers, these were checked for guideline-conformity independently by two of the authors (SK and FS).Criteria for major depression at baseline were assessed with the depression-related modules of the Structured Clinical Interviews for DSM-5 Disorders (SCID-5-CV) [35] two to five days after screening. The interviewers (BSc psychology students) were trained and supervised by the project leader, who is an experienced psychotherapist. Participants who did not meet the criteria for a major depression were considered false positive screens.Depression severity was assessed with the PHQ-9 at one and six months after screening. In accordance with the DSM-5 diagnostic criteria, the PHQ-9 assesses nine depressive symptoms each rated in terms of frequency during the past two weeks (0–3; not at all to nearly every day), resulting in a total score ranging from 0 to 27. The PHQ-9 is among the most frequently used self-report depression questionnaires, has good psychometric properties, and is sensitive to change [25, 36].Suicidal ideation was assessed with the PHQ-9 suicide item (item 9): "Over the last two weeks, how often have you been bothered by thoughts that you would be better off dead or of hurting yourself in some way?", rated from 0–3 (not at all, several days, more than half the days, nearly every day). Emotional response to depressive symptoms was assessed with a composite scale based on two items of the Brief Illness Perception Questionnaire (Brief IPQ) that cover emotional representations of depressive symptoms: "How concerned are you about your symptoms?" and "How much do your symptoms affect you emotionally? (e.g. do they make you angry, scared, upset or depressed)?". The items were assessed directly after the PHQ-9 and were scored on a Likert scale ranging from 0 (not at all) to 10 (extremely). Item scores were pooled for the composite scale, resulting in one total scale ranging from 0 to 10. The respective items of the Brief IPQ showed good psychometric properties [37]."

Outcomes

"Participants were classified as misdiagnosed or mistreated if they reported having received a depression diagnosis by a health professional or guideline-based depression treatment while not having met the criteria for major depression at baseline (SCID), i.e. while being screened false positive. Deterioration in depression severity was defined as a pre-post change score of at least 4.4 points in the PHQ-9. The cut-off is based on the reliable change index (RCI), a psychometric criterion to evaluate whether a change in symptoms is considered statistically reliable, i.e. not attributable to measurement error [38]. The RCI was calculated using the PHQ-9 standard deviation from the current sample (SDbaseline = 4), the reliability coefficient from the PHQ-9 validation study (rtt = 0.84) [39], and a 95% confidence level. Deterioration in emotional response to depressive symptoms was defined as a pre-post change score of at least 3.1 points in the relating composite scale. The RCI was calculated using the standard deviation of this composite scale (SDbaseline = 1.9), the pooled reliability coefficients from the Brief IPQ validation study (rtt = 0.66), and a 95% confidence level.Deterioration in suicidal ideation was defined as the pre-post change score of at least 1 point in the PHQ-9 suicide item."

6a-i) Online questionnaires: describe if they were validated for online use and apply CHERRIES items to describe how the questionnaires were designed/deployed.

The PHQ-9 is preliminarily validated for online use: "Depression was screened as part of the baseline survey using the digitised PHQ-9 [25, 28] (see the outcomes section for further information and supplement A for the layout of the digitised version). At the standard cut-off value of ≥10 points, the paper-pencil PHQ-9 demonstrates high discriminatory performance for detecting major depression: Based on a recent individual participant data meta-analysis of studies with a semi-structured interview reference standard, pooled PHQ-9 sensitivity and specificity (95% confidence interval) were 0.85 (0.79 to 0.89) and 0.85 (0.82 to 0.87), respectively [29]. Preliminary evidence suggests that psychometric characteristics are comparable for the digitised version [30, 31]."

Other outcomes are not validated for online use.

6a-ii) Describe whether and how "use" (including intensity of use/dosage) was defined/measured/monitored

"Of the 787 participants randomised to receive any feedback, in total 744 (95%) opened the feedback screen, of which 464 (62%) downloaded the PDF and 248 (33%) interacted with the feedback by clicking at least one link or modal. There was no descriptive difference between the feedback engagement across feedback arms (see main paper, for results per study arm)."

6a-iii) Describe whether, how, and when qualitative feedback from participants was obtained

Qualitative feedback via interviews was obtained in a separate study.

6b) CONSORT: Any changes to trial outcomes after the trial commenced, with reasons

"We conducted small deviations from the preregistration: we added the outcomes misdiagnosis and emotional response to symptoms, as we deemed this of clinical interest."

7a) CONSORT: How sample size was determined

7a-i) Describe whether and how expected attrition was taken into account when calculating the sample size

Sample size calculation referred to the main analysis. It took into account a dropout rate of 35% and is described in the main paper.

7b) CONSORT: When applicable, explanation of any interim analyses and stopping guidelines

There were no interim analyses.

8a) CONSORT: Method used to generate the random allocation sequence

"After completing baseline assessment and screening, eligible participants were automatically randomised (random permuted blocks randomisation stratified for depression severity generated by a statistician) and allocated 1:1:1 to one of the three study arms. Research staff were masked to allocation at any time until breaking the blind. Due to the design, participants could not be masked but were kept unaware of trial hypotheses to minimise expectancy bias." "

8b) CONSORT: Type of randomisation; details of any restriction (such as blocking and block size)

See 8a).

9) CONSORT: Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned

See 8a).

10) CONSORT: Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions

The randomization sequence was generated by a statistician, uploaded to the platform by the IT specialist, and assigned automatically in order of enrollment.

11a) CONSORT: Blinding - If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing

"Research staff were masked to allocation at any time until breaking the blind. Due to the design, participants could not be masked but were kept unaware of trial hypotheses to minimise expectancy bias."

11a-i) Specify who was blinded, and who wasn't

"Research staff were masked to allocation at any time until breaking the blind. Due to the design, participants could not be masked but were kept unaware of trial hypotheses to minimise expectancy bias."

11a-ii) Discuss e.g., whether participants knew which intervention was the "intervention of interest" and which one was the "comparator"

"Due to the design, participants could not be masked but were kept unaware of trial hypotheses to minimise expectancy bias."

11b) CONSORT: If relevant, description of the similarity of interventions

The content of the nontailored and tailored feedback was broadly similar (see 5-viii).

12a) CONSORT: Statistical methods used to compare groups for primary and secondary outcomes

"We compared the rates of negative effects between study arms in terms of relative risks (RR). The RR estimates how much higher (or lower) the probability of negative effects is for participants in the respective feedback arm compared to the no feedback arm. To directly estimate the RR with 95% confidence intervals, we applied generalised linear models with a log link and robust sandwich variance estimator using modified log-Poisson regressions [41]. We chose this approach over alternative models as it is suited as well in case of frequent outcomes and suffers least from convergence problems [42, 43]."

12a-i) Imputation techniques to deal with attrition / missing values

"Additionally, we performed sensitivity analyses in the intention-to-treat (ITT) sample, both with and without missing data imputation. We used two strategies for imputing data: assuming that all drop-outs were deteriorators, considering this to be the most conservative estimate (worst case); and assuming that all drop-outs were non-deteriorators, considering this to be the most optimistic estimate (best case)."

12b) CONSORT: Methods for additional analyses, such as subgroup analyses and adjusted analyses

"To test for differential effects in the subgroup of false positive screened participants, we ran another series of models additionally including false-positive screens and the false-positive screen x study arm interaction term."

X26) REB/IRB Approval and Ethical Considerations [recommended as subheadingunder "Methods"] (not a CONSORT item)

"Online informed consent via checkboxes was obtained from all participants. The study was approved by the Ethics Committee of the Hamburg Medical Chamber and followed appropriate Consolidated Standards of Reporting Trials (CONSORT) guidelines, including the harms and the e-health statement (see supplement A) [9, 20-24]."

X26-iii) Safety and security procedures

"Due to ethical considerations, all participants who have indicated elevated suicidal ideation (PHQ-9 suicide item ≥ 2; more than half the days) were shown a screen providing an advice to urgently seek help and relevant information on available help services (e.g. general practitioner, local psychiatric emergency units, and the national emergency number; supplement D)."

RESULTS

13a) CONSORT: For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome

See CONSORT flow chart, Figure 2.

13b) CONSORT: For each group, losses and exclusions after randomisation, together with reasons

See CONSORT flow chart, Figure 2.

13b-i) Attrition diagram

See CONSORT flow chart, Figure 2.

14a) CONSORT: Dates defining the periods of recruitment and follow-up

"Recruitment took place from January 2021 to January 2022."

"Data collection was conducted online and in German language between January 12, 2021, and September 30, 2022."

14a-i) Indicate if critical "secular events" fell into the study period

No secular events impacted this study.

14b) CONSORT: Why the trial ended or was stopped (early)

This is not applicable to this study as it was not stopped early.

15) CONSORT: A table showing baseline demographic and clinical characteristics for each group

See Table 1 of the manuscript.

15-i) Report demographics associated with digital divide issues

See Table 1 of the manuscript.

16a) CONSORT: For each group, number of participants (denominator) included in each analysis and whether the analysis was by original

assigned groups

16-i) Report multiple "denominators" and provide definitions

See Table 2.

16-ii) Primary analysis should be intent-to-treat

"We performed this secondary analysis in the per protocol sample which included 948 (88%) out of 1078 randomised participants who had at least one post-baseline value of one of the outcomes and no major protocol violation. The latter were pre-defined as not receiving or adhering to the intervention (i.e., feedback not opened, feedback reading time less than 15 seconds or no download of feedback form), multiple participation (post-hoc data check or self-report), reports of not having answered the survey seriously, baseline survey completion time less than two minutes and provision of an invalid email address. We preferred per protocol over intention-to-treat analysis, as the second is likely to underestimate the risk of an event by inflating the denominator with participants who have provided invalid data or have never received the intervention. Whereas this is conservative in efficacy evaluations, in the current case of a risk evaluation we consider it more appropriate to prevent failing to detect a risk than overestimating it [40]. Additionally, we performed sensitivity analyses in the intention-to-treat (ITT) sample, both with and without missing data imputation. We used two strategies for imputing data: assuming that all drop-outs were deteriorators, considering this to be the most conservative estimate (worst case); and assuming that all drop-outs were non-deteriorators, considering this to be the most optimistic estimate (best case)."

17a) CONSORT: For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95%

confidence interval)

"Misdiagnosis six months after screening was not associated with nontailored (RR=1.3, p=0.509) or tailored feedback (RR=1.09, p=0.843) as compared to no feedback, with rates of 4.9%, 4.1% and 3.5% in the nontailored, the tailored, and the no feedback arm, respectively. Mistreatment six months after screening was not associated with nontailored (RR=0.87, p=0.645) nor tailored feedback (RR=0.95, p=0.859), either, with rates of 7.2%, 7.7%, and 8.3% in the nontailored, the tailored, and the no feedback arm. Descriptively, the rate of mistreatment was higher for psychotherapy (4.9%, 6.4%, and 6.2%) compared to pharmacotherapy (3.4%, 2.2%, and 2.8%). Deterioration in depression severity was not associated with nontailored (one month: RR=1.96, p=0.095; six months: RR=0.6, p=0.143) or tailored feedback (one month: RR=0.7, p=0.494; six

months: RR=0.74, p=0.366), with rates of 5.7%, 2.0%, and 2.9% at one month and 4.1%, 5.1%, and 6.8% at six months in the nontailored, tailored, and no feedback study arm. Deterioration in emotional response to depressive symptoms was not associated with nontailored (one month: RR=1.18, p=0.750; six months: RR=0.46, p=0.197) or tailored feedback (one month: RR=0.23, p=0.128; six months: RR=0.7, p=0.491) either, with rates of 2.7%, 0.7%, and 2.3% at one month and 1.4%, 2%, and 2.9% at six months. Deterioration in suicidal ideation was not associated with nontailored (RR=1.12, p=0.655) or tailored feedback (RR=1.4, p=0.147) at six months, with rates of 10.5%, 13.1%, and 9.4%. At one month, it was almost two-fold increased in the nontailored (RR=1.92; p=0.014), but not in the tailored feedback arm (RR=1.26, p=0.427), as compared to no feedback. Rates in the nontailored, the tailored, and the no feedback arm were 12.3%, 8.1%, and 6.4%. Absolute numbers and rates for all negative effects per study arm and time point are shown in Table 2. Relative risks with corresponding 95% confidence intervals are illustrated in Figure 3."

17a-i) Presentation of process outcomes such as metrics of use and intensity of use

There were no process outcomes assessed.

17b) CONSORT: For binary outcomes, presentation of both absolute and relative effect sizes is recommended

See Table 2 of the manuscirpt.

18) CONSORT: Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory

"Results did not differ for the subgroup of false-positives (Pinteraction ranging between 0.287 and 0.804). Sensitivity analyses based on logistic regression models as well as those in the ITT sample with the full analysis set and with missing data imputation based on the best case scenario showed comparable results. In the ITT analysis based on the worst case scenario, however, the relative risk for deterioration in suicidal ideation in the nontailored feedback arm at one month was not higher than in the no feedback arm (RR=1.26, p=0.065; supplement D). Post hoc analyses exploring baseline demographic and clinical characteristics of all participants deteriorated in any outcome at any time point were comparable to the total sample (supplement F)."

18-i) Subgroup analysis of comparing only users

As mentioned, the per protocol analysis was our main analysis, as this seemd mre appropriate in the context of negative effects (see above).

19) CONSORT: All important harms or unintended effects in each group

All results relate to negative effects.

19-i) Include privacy breaches, technical problems

There were no privacy breaches or unexpected/unintended incidents.

19-ii) Include qualitative feedback from participants or observations from staff/researchers

No qualitative feedback was collected in this study.

DISCUSSION

20) CONSORT: Trial limitations, addressing sources of potential bias, imprecision, multiplicity of analyses

20-i) Typical limitations in ehealth trials

"The interpretation of these results should be considered in the context of the study's limitations. First, the underlying DISCOVER trial did not explicitly call for those seeking depression screening. As these may be more eager to follow the advice of the feedback, in this sample misdiagnosis and mistreatment might be underestimated. A second limitation is that due to the design of the DISCOVER trial, the selection of outcomes was limited and relevant negative effects such as distress, stigma, treatment side effects, or overdiagnosis (i.e. the diagnosis of correctly diagnosed but mild cases that would not benefit from treatment [14]) could not be assessed. Third, all outcomes were self-reported. Although this is common in psychological interventions, particularly the assessment of misdiagnosis and mistreatment would benefit from more objetive data from

health care providers. Fourth, the operationalisations of suicidal ideation and emotional response to depressive symptoms are based on a single item and a composite score not well validated for this purpose. Indeed, evidence for the validity of the PHQ-9 suicide item is inconclusive, with studies indicating both good prediction versus overestimation of suicidal ideation or attempts [44, 45]. Lastly, the study was planned post-hoc and therefore not powered to detect the selected outcomes, and multiple testing might have led to overestimation of significance in the case of deterioration in suicidal ideation. Notably, the findings refer to the German health care system where psychotherapy is available and covered by the social health insurance. Particularly rates for misdiagnosis

21) CONSORT: Generalisability (external validity, applicability) of the trial findings

21-i) Generalizability to other populations

See 20-i).

21-ii) Discuss if there were elements in the RCT that would be different in a routine application setting

This intervention was delivered as design for practice.

22) CONSORT: Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence

22-i) Restate study questions and summarize the answers suggested by the data, starting with primary outcomes and process outcomes (use)

"To the best of our knowledge, this secondary analysis is the first study to systematically examine potential negative effects of feedback after internet-based depression screening in a large sample of currently undiagnosed and untreated individuals with at least moderate depression severity. The results indicate that feedback, both nontailored and tailored, was not associated with increased rates of misdiagnosis, mistreatment, deterioration in depressive symptoms, or deterioration in emotional response to symptoms as compared to no feedback. Deterioration of suicidal ideation, however, appeared to be more likely one month after receiving nontailored feedback compared with no feedback; an association that was not found any more at six-months follow-up and neither after tailored feedback. Although almost 40% of the sample turned out to be screened false positive, irrespective of the study arm rates of subsequent misdiagnosis and mistreatment were lower than 5% and 9%, respectively, with rates of pharmacotherapy ranging even lower than 4%. Across study arms, deterioration in emotional response to depressive symptoms was reported by at most 3% of participants, deterioration of depression severity by at most 7%, and deterioration of suicidal ideation by at most 13%."

22-ii) Highlight unanswered new questions, suggest future research

"Therefore, comparing outcomes such as distress or negative affectivity shortly after providing the screening vs. the feedback appears worthwile to further address these issues (see [48, 49] for examplary study designs in suicide screening)."

"Given that these results should be interpreted with caution due to the study's limitations, more robust research is needed to further address suicidal ideation in internet-based depression screening. If prospective trials that use validated ouctome measures corroborate an association of internet-based screening and/or feedback with suicidal ideation, this should inform regulations of currently unmonitored internet-based depression tests. Further, the findings should also inform research regarding comparable depression screening in medical and primary care settings, which is currently recommended in many countries despite very uncertain evidence regarding potential harms [51]."

OTHER INFORMATION

23) CONSORT: Registration number and name of trial registry

Trial Registration: ClinicalTrials.gov (NCT04633096)
Preregistration of secondary data analysis: OSF.io (https://osf.io/tzyrd)

24) CONSORT: Where the full trial protocol can be accessed, if available

Sikorski, F., et al., The efficacy of automated feedback after internet-based depression screening: Study protocol of the German, three-armed, randomised controlled trial DISCOVER. Internet Interventions, 2021. 25: p. 100435.

25) CONSORT: Sources of funding and other support (such as supply of drugs), role of funders

"This work was funded by the German Research Foundation as part of the underlying DISCOVER RCT (grant number: 424162019)."

X27-i) State the relation of the study team towards the system being evaluated

In terms of conflicts, "None declared".

Multimedia Appendix 2

Multimedia Appendix: Illustration of the digitised PHQ-9 as displayed to study participants (mobile version).

Multimedia Appendix 6: Characteristics of the ITT sample

Table. Baseline demographic and clinical characteristics of the intention-to-treat sample (*N* = 1178).

	No feedback (*n* = 391)	Non-tailored feedback (*n* = 393)	Tailored feedback (*n* = 394)
Age, years	36.5 (13.8)	37.7 (14.0)	37.2 (14.8)
Gender			
Female	276 (71%)	275 (70%)	273 (69%)
Male	111 (28%)	115 (30%)	118 (30%)
Diverse	4 (1%)	3 (1%)	3 (1%)
German mother tongue	369 (94%)	370 (94%)	379 (96%)
Migration background	47 (12%)	35 (9%)	39 (10%)
Being in a relationship	167 (43%)	200 (51%)	192 (49%)
Living together	258 (66%)	277 (71%)	265 (67%)
Formal school education			
Low (less than 10 years)	71 (18%)	80 (20%)	68 (17%)
Middle (at least 10 years)	120 (31%)	129 (33%)	134 (34%)
High (A-level or above)	200 (51%)	184 (47%)	192 (49%)
Working	276 (71%)	278 (71%)	293 (74%)
Quality of life (EQ-5D-5L VAS)	57.7 (22.4)	56.8 (22.2)	58.6 (22.0)
Depression severity (PHQ-9)	14.8 (4.0)	14.8 (4.1)	14.7 (3.9)
Emotional response	6.9 (2.1)	6.9 (2)	7 (1.8)
Anxiety severity (GAD-7)	12.0 (4.3)	12.3 (4.3)	11.9 (4.3)
Somatic symptom severity (SSS-8)	14.5 (5.3)	14.5 (5.1)	14.4 (5.3)
No. of depression-related risk factors[a]	6.0 (2.5)	6.1 (2.4)	5.8 (2.3)
Frequency of suicidal ideation within last two weeks			
None	167 (51%)	161 (51%)	165 (54%)
Several days	98 (30%)	113 (36%)	94 (31%)
More than half the days	37 (11%)	23 (7%)	26 (9%)
Nearly every day	25 (8%)	17 (5%)	22 (7%)
Self-identifying as suffering from depression			
No	55 (14%)	44 (11%)	66 (17%)
Maybe	162 (41%)	201 (51%)	176 (45%)
Yes	174 (45%)	148 (38%)	152 (39%)
Meeting criteria for major depression (SCID)	194 (62%)[b]	180 (61%)[c]	180 (60%)[d]

Data are mean (SD) or n (%). PHQ-9=Patient Health Questionnaire-9. EQ-5D-5L=EuroQoL-5 Dimensions-5 Level scale. VAS=visual analogue scale. GAD-7=Generalized Anxiety Disorder-7. SSS-8=Somatic Symptom Scale. [a]Risk factors included self-reported anxiety, addiction, traumatic life events, persistent physical symptoms, mood swings, chronic physical condition, lack of social support, mental comorbidity, mental comorbidity in family, history of suicide, current pregnancy, post-natal phase, menopause, premenstrual syndrome. SCID=Structured Clinical Interview for DSM-5 Disorders; the interview was conducted approximately 2 to 5 days after randomisation. [b]78 cases with missing data. [c]97 cases with missing data. [d]94 cases with missing data.

Multimedia Appendix 7: Sensitivity analyses

Table 7.1. Rates and relative risks of negative effects per study arm and time point in the intention-to-treat sample in full analysis set.

	N	NF	N	NTF	N	TF	Relative risk (95% CI)	
							NTF vs. NF	TF vs. NF
Misdiagnosis	313[a]	11 (3.5%)	296[a]	13 (4.4%)	300[a]	11 (3.7%)	1.25 [0.57 – 2.75]	1.04 [0.46 – 2.37]
Mistreatment	313[a]	24 (7.7%)	296[a]	19 (6.4%)	300[a]	21 (7%)	0.84 [0.47 – 1.5]	0.91 [0.52 – 1.6]
PSYCHOTHERAPY	313[a]	18 (5.8%)	296[a]	13 (4.4%)	300[a]	17 (5.7%)	0.76 [0.38 – 1.53]	0.99 [0.52 – 1.88]
PHARMACOTHERAPY	313[a]	8 (2.6%)	296[a]	9 (3%)	300[a]	6 (2%)	1.19 [0.47 – 3.04]	0.78 [0.28 – 2.23]
Deterioration in depression								
1 month	329	11 (3.3%)	325	18 (5.5%)	322	7 (2.2%)	1.66 [0.8 – 3.45]	0.65 [0.26 – 1.66]
6 months	325	22 (6.8%)	319	14 (4.4%)	321	18 (5.6%)	0.65 [0,34 – 1.24]	0,83 [0.45 – 1.52]
Deterioration in suicidality								
1 month	329	25 (7.6%)	325	42 (12.9%)	322	28 (8.7%)	1.7 [1.06 – 2.72]	1.14 [0.68 – 1.92]
6 months	325	33 (10.2%)	319	34 (10.7%)	321	44 (13.7%)	1.05 [0.67 – 1.65]	1.35 [0.88 – 2.06]
Deterioration in emotional response								
1 month	325	9 (2.8%)	323	9 (2.8%)	320	3 (0.9%)	1.06 [0.41 – 2.76]	0.32 [0.07 – 1.48]
6 months	322	9 (2.8%)	317	6 (1.9%)	317	8 (2.5%)	.44 [0.14 – 1.37]	0.76 [0.29 – 2]

Data are n (%). NF=No feedback arm. NTF=Non-tailored feedback arm. TF=Tailored feedback arm. CI=Confidence Interval.

Table 7.2. Rates and relative risks of negative effects per study arm and time point in the intention-to-treat sample with missing data imputation; missing data = 0 – best case.

							Relative risk (95% CI)	
	N	NF	N	NTF	N	TF	NTF vs. NF	TF vs. NF
Misdiagnosis	391	11 (2.8%)	393	13 (3.3%)	394	11 (2.8%)	1.18 [0.53 – 2.59]	0.99 [0.44 – 2.26]
Mistreatment	391	24 (6.1%)	393	19 (4.8%)	394	21 (5.3%)	0.79 [0.44 – 1.41]	0.87 [0.49 – 1.53]
PSYCHOTHERAPY	391	18 (4.6%)	393	13 (3.3%)	394	17 (4.3%)	0.72 [0.36 – 1.45]	0.94 [0.49 – 1.79]
PHARMACOTHERAPY	391	8 (2%)	393	9 (2.3%)	394	6 (1.5%)	1.12 [0.44 – 2.87]	0.74 [0.26 – 2.13]
Deterioration in depression								
1 month	391	11 (2.8%)	393	18 (4.6%)	394	7 (1.8%)	1.63 [0.78 – 3.4]	0.63 [0.25 – 1.61]
6 months	391	22 (5.6%)	393	14 (3.6%)	394	18 (4.6%)	0.63 [0.33 – 1.22]	0.81 [0.44 – 1.5]
Deterioration in suicidality								
1 month	391	25 (6.4%)	393	42 (10.7%)	394	28 (7.1%)	1.67 [1.04 – 2.67]	1.11 [0.66 – 1.87]
6 months	391	33 (8.4%)	393	34 (8.7%)	394	44 (11.2%)	1.03 [0.65 – 1.62]	1.32 [0.86 – 2.03]
Deterioration in emotional response								
1 month	391	9 (2.3%)	393	9 (2.3%)	394	3 (0.8%)	1 [0.4 – 2.48]	0.33 [0.9 – 1.21]
6 months	391	9 (2.3%)	393	6 (1.5%)	394	8 (2%)	0.66 [0.24 – 1.85]	0.88 [0.34 – 2.26]

Data are n (%). NF=No feedback arm. NTF=Non-tailored feedback arm. TF=Tailored feedback arm. CI=Confidence Interval.

Table 7.3. Rates and relative risks of negative effects per study arm and time point in intention-to-treat sample with missing data imputation; missing data = 1 – worst case.

							Relative risk (95% CI)	
	N	NF	N	NTF	N	TF	NTF vs. NF	TF vs. NF
Misdiagnosis	391	98 (25.1%)	393	116 (29.5%)	394	112 (28.4%)	1.18 [0.94 – 1.48]	1.13 [0.9 – 1.43]
Mistreatment	391	102 (26%)	393	116 (29.5%)	394	116 (29.4%)	1.13 [0.9 – 1.42]	1.13 [0.9 – 1.42]
PSYCHOTHERAPY	391	96 (24.6%)	393	110 (28%)	394	111 (28.2%)	1.14 [0.9 - 1.44]	1.15 [0.91 – 1.45]
PHARMACOTHERAPY	391	86 (22%)	393	106 (27%)	394	100 (25.4%)	1.23 [0.96 – 1.57]	1.15 [0.9 – 1.49]
Deterioration in depression								
1 month	391	73 (18.7%)	393	86 (21.9%)	394	79 (20.1%)	1.17 [0.89 – 1.55]	1.07 [0.81 – 1.43]
6 months	391	88 (22.5%)	393	88 (22.4%)	394	91 (23.1%)	1 [0.77 – 1.29]	1.03 [0.79 – 1.33]
Deterioration in suicidality								
1 month	391	87 (22.3%)	393	110 (28%)	394	100 (25.4%)	1.26 [0.99 – 1.61]	1.14 [0.89 – 1.47]
6 months	391	99 (25.3%)	393	108 (27.5%)	394	117 (29.7%)	1.01 [0.86 – 1.37]	1.17 [0.93 – 1.47]
Deterioration in emotional response								
1 month	391	75 (19.2%)	393	79 (20.1%)	394	77 (19.5%)	0.69 [0.32 – 1.52]	0.45 [0.17 – 1.18]
6 months	391	78 (19.9%)	393	82 (20.9%)	394	85 (21.6%)	1.05 [0.79 – 1.38]	1.08 [0.82 – 1.42]

Data are n (%). NF=No feedback arm. NTF=Non-tailored feedback arm. TF=Tailored feedback arm. CI=Confidence Interval. [a]n refers to participants who completed both the follow-up assessment and the SCID depression module at baseline.

Table 7.4. Rates and odds ratios of negative effects per study arm and time point in per protocol sample, based on logistic regression.

							Odds ratios (95% CI)	
	N	NF	N	NTF	N	TF	NTF vs. NF	TF vs. NF
Misdiagnosis	290[a]	11 (3.5%)	263[a]	13 (4.9%)	267[a]	11 (4.1.%)	1.31 [0.58 – 3]	1.09 [0.47 – 2.56]
Mistreatment	290[a]	24 (8.3%)	263[a]	19 (7.2%)	267[a]	21 (7.7%)	0.86 [0.46 – 1.62]	0.95 [0.51 - 1.74]
Psychotherapy	290[a]	18 (6.2%)	263[a]	13 (4.9%)	267[a]	17 (6.4%)	0.79 [0.38 – 1.64]	1.03 [0.52 – 2.04]
Pharmacotherapy	290[a]	8 (2.8%)	263[a]	9 (3.4%)	267[a]	6 (2.2%)	1.25 [0.48 – 3.27]	0.81 [0.28 – 2.37]
Deterioration in depression								
1 month	312	9 (2.9%)	300	17 (5.7%)	297	6 (2.0%)	2.02 [0.89 – 4.61]	0.69 [0.24 – 1.98]
6 months	309	21 (6.8%)	296	12 (4.1%)	297	15 (5.1%)	0.58 [0,28 – 1.2]	0,73 [0.37 – 1.44]
Deterioration in suicidality								
1 month	312	20 (6.4 %)	300	37 (12.3%)	297	24 (8.1%)	2.05 [1.16 – 3.63]**	1.28 [0.69 – 2.38]
6 months	309	29 (9.4%)	296	31 (10.5%)	297	39 (13.1%)	1.13 [0.66 – 1.9]	1.46 [0.88 – 2.43]
Deterioration in emotional response								
1 month	308	7 (2.3%)	299	8 (2.7%)	296	2 (0.7%)	1.18 [0.42 – 3.3]	0.29 [0.06 – 1.42]
6 months	307	9 (2.9%)	294	4 (1.4%)	299	6 (2%)	0.46 [0.14 – 1.5]	0.69 [0.24 – 1.97]

Data are n (%). NF=No feedback arm. NTF=Non-tailored feedback arm. TF=Tailored feedback arm. CI=Confidence Interval. [a]n refers to participants who completed both the follow-up assessment and the SCID depression module at baseline.

Multimedia Appendix 8: Post hoc analyses

Table. Baseline and clinical characteristics of deteriorators (in any outcome) in the per protocol sample (*N* = 203).

Age, years	37.4 (14.6)
Gender	
Female	138 (68%)
Male	63 (31%)
Diverse	2 (1%)
German mother tongue	197 (97%)
Migration background	17 (9%)
Being in a relationship	104 (51%)
Living together	139 (69%)
Formal school education	
Low (less than 10 years)	44 (22%)
Middle (at least 10 years)	67 (33%)
High (A-level or above)	92 (45%)
Working	131 (65%)
Quality of life (EQ-5D-5L VAS)	56.2 (23.7)
Depression severity (PHQ-9)	14.3 (3.6)
Emotional response to depressive symptoms (composite scale)	6.8 (2)
Anxiety severity (GAD-7)	12 (4.5)
Somatic symptom severity (SSS-8)	14.5 (5.2)
No. of depression-related risk factors[a]	6.2 (2.4)
Frequency of suicidal ideation within last two weeks (PHQ-9 item 9)	
None	125 (62%)
Several days	61 (30%)
More than half the days	16 (8%)
Nearly every day	1 (0.5%)
Self-identifying as suffering from depression	
No	23 (11%)
Maybe	98 (48%)
Yes	82 (40%)
Meeting criteria for major depression (SCID)	115 (70%)[b]

Data are mean (SD) or n (%). PHQ-9=Patient Health Questionnaire-9. EQ-5D-5L=EuroQoL-5 Dimensions-5 Level scale. VAS=visual analogue scale. GAD-7=Generalized Anxiety Disorder-7. SSS-8=Somatic Symptom Scale. [a]Risk factors included self-reported anxiety, addiction, traumatic life events, persistent physical symptoms, mood swings, chronic physical condition, lack of social support, mental comorbidity, mental comorbidity in family, history of suicide, current pregnancy, post-natal phase, menopause, premenstrual syndrome. SCID=Structured Clinical Interview for DSM-5 Disorders; the interview was conducted approximately 2 to 5 days after randomisation. [b]38 cases with missing data.

7.4. Publication IV

How adults with suspected depressive disorder experience online depression screening: A qualitative interview study

Sikorski, F., Löwe, B., & Kohlmann, S.

Status: Published in Internet Interventions online, October 23, 2023.

Internet Interventions 34 (2023) 100685

Contents lists available at ScienceDirect

Internet Interventions

journal homepage: www.elsevier.com/locate/invent

How adults with suspected depressive disorder experience online depression screening: A qualitative interview study

Franziska Sikorski *, Bernd Löwe, Sebastian Kohlmann

Department of Psychosomatic Medicine and Psychotherapy, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

ARTICLE INFO

Keywords:
Depression
Online screening
Early detection
Reflexive thematic analysis
Patient perspective
Qualitative study

ABSTRACT

Background: While evidence on the effects and mechanisms of online depression screening is inconclusive, publicly available 'online depression tests' are already frequently used. To further a comprehensive understanding of online depression screening and evince the perspectives of those affected, we aimed to qualitatively explore how adults with undiagnosed but suspected depressive disorder experience the screening process.
Methods: This study is a qualitative follow-up of a German-wide, 3-arm, randomised controlled trial on feedback after online depression screening conducted between Jan 2021 and Sep 2022. A subsample of 26 participants with undiagnosed but suspected depressive disorder (Patient Health Questionnaire-9 $\geq$ 10; no depression diagnosis/treatment within the last year) were purposefully selected based on maximum variation in gender, age, and study arm. In-depth semi-structured telephone interviews (mean = 37 min) were conducted approximately six months after screening. Data were analysed within a contextualist theoretical framework using inductive reflexive thematic analysis.
Results: Participants were balanced in terms of gender (female/male, n = 15/11), age (range = 22 to 61 years), and study arm (no feedback/standard feedback/tailored feedback, n = 7/11/8). Reported experiences of online depression screening can be described as a two-step process: Step 1 is the initial reaction to the screening procedure and comprises the theme *recognition of depressive symptoms: from denial to awareness*. Step 2 describes a subsequent self-explorative process encompassing the themes *cognitive positioning: rejection* vs. *acceptance, emotional reaction: between overload and empowerment,* and *personal activation: from reflection to action.*
Conclusions: Findings indicate that online depression screening with and without feedback of results is experienced as a two-step process promoting symptom recognition and subsequent self-exploration. While few participants reported negative effects, the majority described the screening process as insightful, empowering, and activating. Future research should determine to what extent online depression screening may pose a standalone form of low-threshold support for individuals with undiagnosed depressive disorder, while focusing as well on potential negative effects.

1. Background

Major depression is one of the most disabling and most prevalent disorders worldwide (GBD 2019 Mental Disorders Collaborators, 2022). Yet, affected individuals still often go undetected: In primary care, for example, only 50 % of depressed patients are correctly diagnosed and treated (Mitchell et al., 2009; Trautmann and Beesdo-Baum, 2017), and patients who eventually make a treatment contact do so with an average delay of eight years after depression onset (Wang et al., 2005). Without treatment, however, depressive symptoms can worsen over time, resulting in an increased likelihood of a chronic course, a worse treatment outcome, rising healthcare costs, and an increased disease burden (Kraus et al., 2019).

While traditional service uptake is low, individuals increasingly seek mental health information on the internet (Berger et al., 2005; Eichenberg et al., 2013), with the use of online depression screening being on the rise. In 2020, for example, nearly 2.6 million online mental health screeners were completed through the website of only one American mental health organisation (Mental Health America; Kruzan et al., 2022) - which joins a multitude of other health-related platforms and apps that

* Corresponding author at: Department of Psychosomatic Medicine and Psychotherapy, University Medical Center Hamburg-Eppendorf, Martinistr. 52, 20246 Hamburg, Germany.
E-mail address: f.sikorski@uke.de (F. Sikorski).

https://doi.org/10.1016/j.invent.2023.100685
Received 11 May 2023; Received in revised form 9 October 2023; Accepted 21 October 2023
Available online 23 October 2023

provide publicly available online depression screening. The so called 'online depression tests' typically use self-report measures of depression symptom severity (e.g., the Patient Health Questionnaire-9) and then provide individuals with feedback on their results, sometimes supplemented by links or referrals to services. Aiming at empowering individuals to better understand and act on their symptoms, depression screening with feedback may provide an important form of support to affected individuals and is proposed to be a promising approach to promote early detection and subsequent resolution of undiagnosed depression (Hassem, 2022; Kohlmann et al., 2020; Kruzan et al., 2022; Löwe et al., 2016; Sikorski et al., 2021).

Despite growing public use and recognition of online depression screening as a promising way of early detection, evidence in this field is limited and inconsistent. With regard to psychometric validity for example, a systematic review identified that the screening accuracy of thirteen examined online depression screeners varied significantly across different samples as well as between and within conditions and instruments (Martin-Key et al., 2022). Few studies also addressed the screening efficacy, i.e. the actual merit of the resulting diagnostic information to patient-related outcomes such as help-seeking. In one observational study on users of a depression screening app who received feedback on their results, approximately 38 % of participants reported to have consulted a health professional after one month (BinDhim et al., 2016). Another study investigated online search behaviour after completion of online mental health screening and found that individuals who underwent online depression screening were more likely to conduct subsequent depression-related online searches (Jacobson et al., 2022). More rigorous research, however, has failed to confirm positive effects on help-seeking: In the only published randomised controlled trial on online depression screening, feedback (vs. no feedback) of screening results had no significant effect on professional help-seeking three months later (Batterham et al., 2016).

In addition to providing an only inconclusive picture of the effects and potential mechanisms of online depression screening, this quantitative evidence further omits the voices and perspectives of those affected. Qualitative research, by contrast, has the potential to both complement the understanding of the complexity of online depression screening and provide insights into the life-worlds and subjective health needs of affected individuals using it. However, so far only one study has addressed the individuals' perspectives in this matter: By conducting focus groups with young adults who voluntarily sought a screening website, this study showed that online depression screening met young adults' emotional needs for validation and self-understanding. It further suggested that online screening can serve as a transition point in young people's mental health journeys (Kruzan et al., 2022). Although this study expands on quantitative findings by highlighting the potential value of the screening process itself, it is restricted to a very young population and did not exclude cases already diagnosed and/or in care.

In the present study, we seek to further a comprehensive understanding of online depression screening by exploring the perspectives of adults of all ages who are undiagnosed but screened positive for at least moderate depressive symptomatology within a randomised controlled trial. Specifically, we aim to better understand how adults with undiagnosed but suspected depressive disorder experience the screening process.

2. Methods

2.1. Study context and design

The study was conducted as a qualitative follow-up of a randomised controlled trial (RCT) on feedback after online depression screening, conducted nationwide in Germany between January 2021 and September 2022 (see Sikorski et al., 2021, for the study protocol; main manuscript under preparation). After completing online depression screening with the Patient Health Questionnaire-9 (PHQ-9, Kroenke et al., 2001), a total of 1178 participants with undiagnosed but suspected depressive disorder (PHQ-9 $\geq$ 10) were randomised to get either no feedback (n = 391), standard feedback (n = 393), or tailored feedback (n = 394) on the screening result. Online assessments were conducted at baseline, 1-month, and 6-months follow-up and were complemented by diagnostic telephone interviews (SCID) at baseline and 6-months follow-up.

This qualitative study is based on a purposefully selected subsample of participants who were interviewed following the 6-months follow-up assessment. The research question was addressed within a paradigmatic framework of contextualism, assuming that observable data is informative of an existing reality, but does not straightforwardly reflect it. In alignment with an explorative design and in order to capture participants' experiences as close to their own accounts as possible, data analysis was approached inductively.

2.2. Study sample

Participants of the RCT were required to be aged 18 years or above, screen positive for suspected depressive disorder (PHQ-9 $\geq$ 10), provide contact details, have sufficient German language as well as computer literacy, have internet access, and be willing to give informed consent. Participants were excluded if they reported to have been diagnosed with or treated for depression within the past 12 months.

The present subsample was purposefully selected, adopting maximum variation sampling to strive for an in-depth understanding across a wide range of perspectives rather than empirical generalisability (Palinkas et al., 2015). Individuals were selected based on variation in the following primary sampling criteria: gender (male, female), study arm (no feedback, standard feedback, tailored feedback), and age (<40 years, ≥40 years), and, if feasible, also variation in reported depression history (depression diagnosis in the past yes vs. no; self-report) and depression severity at time of screening (moderate, severe; PHQ-9). To reduce the probability of possible bias by extreme cases, the aim was to recruit two participants per combination of primary sampling criteria (age, gender, study arm), i.e. 24 participants in total.

2.3. Recruitment

The RCT was promoted nationwide as a 'German-wide study on stress and psychological well-being' (www.discover-studie.de). The aim of evaluating online depression screening was not explicitly communicated, but interested participants were informed that some of them will get feedback on a part of their answers. Participants were recruited from the general population through traditional and social media, print advertisement in public areas, and a population wide online access survey panel to strive for a balanced composition of the sample (January 2021 to February 2022). Participation was compensated with vouchers worth up to 15 euros.

Recruitment for the qualitative study was conducted on an ongoing basis at the end of the 6-months follow-up interview of the RCT (July 2021 to August 2022). In this context, the study was presented to participants as being part of the first author's (FS) PhD project. Out of 1075 interviewed participants, 806 gave oral consent to be contacted for the qualitative study. Subsequently, 135 participants who met maximum variation sampling criteria were provided with detailed study information via email. Of those, 26 participants returned electronic or written informed consent and were scheduled for an interview appointment. Participation was compensated with 10 euros (vouchers).

2.4. Online depression screening and feedback

The PHQ-9 (Kroenke et al., 2001; German translation: Löwe et al., 2002) is a widely used and easily administered depression screening tool. For the recommended cut off point of 10, it demonstrates robust

psychometric characteristics and a high discriminatory performance for detecting a major depression in both the paper-pencil and the online version (Du et al., 2017; Erbe et al., 2016; Miller et al., 2021). It consists of nine items covering all major depression symptom criteria as stated in the DSM-5 ('Over the past two weeks, how often have you been bothered by any of the following problems?'). Each item is scored on a 4-point Likert scale ranging from 'not at all' (0) to 'nearly every day' (3), resulting in a total score ranging from 0 to 27 with scores of 10 and 15 indicating moderate and severe depressive symptoms.

The PHQ-9 was embedded in a baseline survey comprising additional questions on personal data, sociodemographic characteristics, and other health-related outcomes (e.g., depression-related illness beliefs). Participants who indicated elevated suicidal ideation (PHQ-9 suicide item $\geq$2) were directly shown a screen providing an advice to urgently seek help and relevant information on available help services (e.g. general practitioner, local psychiatric emergency units, and the national emergency number). After completing the survey, all randomised participants were thanked for participating in the study and received information on follow-up procedures.

In case participants received feedback, it consisted of (1) the depression screening result, (2) a note to seek diagnostic consultation by a health professional, (3) brief general information on depression, and (4) information on depression treatment with direct links to referenced health or social services (see the study website for a German demo). In extension to the standard version, the tailored feedback was personalised to participants' characteristics as follows: by phrasing screening result (1) and general information on depression (3) according to participants' symptom profiles and indicated causal attributions (e.g., 'You have indicated that you had *low spirits*, *sleep disturbances*, and *loss of energy* during the past two weeks.'), by matching the note to seek further consultation (2) to participants' specialist preferences (general practitioner vs. mental health professional), and by adapting help seeking advices (4) to participants' health insurance provider and local residency (e.g. by providing links to self-help groups located nearby). Additionally, after being provided with the screening result (1), participants were asked whether they think that their symptoms were indications of depression and whether they worried about the symptoms. According to participants' answers, the following three feedback sections were arranged in a differing order and were phrased slightly differently (see Fig. 1 and supplemental Fig. III in Sikorski et al., 2021, for examples).

2.5. Data collection

A semi-structured interview guide was developed to structure qualitative data collection. Initial questions on motivation for participation in the RCT and symptoms experienced at that time aimed at helping participants to recall the screening situation. Subsequent questions focused on the experience of screening questions or feedback, related health behaviour, an evaluation of the feedback provided, and attitudes towards online depression tests in general (see Supplementary Table 2). The interview guide was discussed in a doctoral colloquium on qualitative research and was piloted within the research team and with the first study participant, resulting in small modifications. Demographic and clinical characteristics of participants were obtained from the RCT (see Sikorski et al., 2021).

Interviews were conducted via telephone from July 2021 to August 2022 by FS, with two interviews each accompanied by another study team member. Probes and clarifying questions were used to encourage participants to elaborate on their experiences and to express both positive and negative accounts in order to reduce possible bias. Due to the explorative nature of the research question, discussions were also guided by what FS interpreted to be meaningful to the interviewee. Interviews were audiotaped, pseudonymised and transcribed verbatim by trained student research assistants (MSc Psychology candidates). Transcription followed the rules of Dresing and Pehl (2015), with all transcripts being checked for correctness by FS. Interviews took place on average 211 days after screening ($SD = 20.9$) and the mean length was 37 min (range: 15 min to 1 h 14 min).

2.6. Data analysis

Data were analysed using reflexive thematic analysis, a theoretically flexible and interpretative approach to identify themes within and between participants' accounts in qualitative data (Braun and Clarke, 2006, 2019, 2021). In line with the contextualist paradigm, analysis was approached through a critical realist epistemological perspective, i.e. assuming the existence of an external reality, but acknowledging that the way individuals make meaning of their experience and therefore access to knowledge is socially influenced. Data interpretation followed an experiential orientation, i.e. examining accounts and meaningfulness as ascribed by participants. For coding, a research question-led, inductive approach with both semantic and latent coding was adopted. The analytic process followed Braun and Clarke's (2006) six-phase process: (1) Familiarisation with the data was done by re-listening and -reading all interviews and by taking notes on first impressions. (2) Coding as well as (3) developing, (4) reviewing and (5) naming of themes were conducted in an organic, iterative and recursive process. In line with our research design, themes were developed by clustering codes around a 'central organising concept' (Braun and Clarke, 2019) drawing on meaningfulness rather than frequency of mentions as a central criterion. The process concluded with (6) selecting appropriate quotations and producing the report.

Data analysis was conducted using the software MAXQDA (version 2022) and was led by FS and supervised by the last author SK (August 2022 to February 2023). Both authors met regularly to reflect on potential pre-assumptions, interpretations of codes, and theme development to achieve reflexive engagement with data and ultimately agreement on themes. Translation of cited quotations from German to English language considered the transfer of meaning, sense, and context and was conducted by FS, followed by a final discussion with SK. For the report, some quotations were edited for brevity purposes (indicated by […]) and grammatical and spelling errors were corrected to facilitate readability and comprehension. Quotations are marked with a corresponding participant number, gender, age range, and study arm.

2.7. Researcher statement

FS is a female clinical psychologist conducting a psychodynamic psychotherapy training and a PhD training programme, in which she is attending monthly colloquia on qualitative research. SK (PhD, CBT psychotherapist) and BL (MD, CBT and psychodynamic psychotherapist) are both senior researchers experienced with both quantitative and qualitative research on depression.

2.8. Ethics and good clinical practice

The study is designed and reported according to the COREQ and the JARS-QUAL guidelines for qualitative research (Levitt et al., 2018; Tong et al., 2007; see Supplementary Table 1 for the filled COREQ checklist) and specific guidelines for promoting more deliberate and reflexive engagement in thematic analysis research (Braun and Clarke, 2021). All procedures involved in the study have been approved by the Ethics

Committee of the University Medical Center Hamburg-Eppendorf (June 2021, reference: 0337).

3. Results

3.1. Participant characteristics

Maximum variation sampling was achieved, with only one under-recruited combination due to little response (young males without feedback, $n = 1$) and one over-recruited combination due to mis-categorisation (young females with standard feedback, $n = 5$). The resulting subsample of 26 participants was balanced in terms of gender (female/male, $n = 15/11$), study arm (no/ standard/ tailored feedback, $n = 7/11/8$), and age strata ($<40/ \geq 40$ years, $n = 14/12$). Age ranged from 22 to 61 years with a mean of 48.8 years ($SD = 12.9$). At time of screening, participants reported on average severe depressive symptoms (PHQ-9, $M = 15.4$, $SD = 4.77$), with 13 participants each displaying moderate and severe depressive symptoms. More than half of the participants did not have any depression diagnosis in the past ($n = 14$). Most participants were in a relationship ($n = 21$) and cohabiting ($n = 20$). Most participants worked ($n = 19$) and about two third reported a high-educated level ($n = 17$; International standard classification of education [ISCED], UNESCO Institute for Statistics, 2012). The 19 participants from the feedback study arms spent on average 13 min on the feedback screen ($SD = 28$), with 12 participants reporting to remember having received feedback at the 6-months follow-up. None of the participants reported negative effects attributed to screening or study participation. Selected characteristics per participant are presented in Supplementary Table 3.

3.2. Themes

Most participants offered diverse accounts of how they experienced online depression screening, which we organised into four themes. As illustrated in Fig. 1, we found these themes to follow a two-step process: Step 1 is the initial reaction to the screening procedure and comprises the theme *recognition of depressive symptoms*: *from denial to awareness*. Step 2 describes a subsequent self-explorative process that encompassed up to three themes: *cognitive positioning* describes the participants' reports on how they related to an illness-related self-identity in reaction to the screening. *Emotional reactions* reported by the participants were often ambivalent and ranged between the poles of overload and empowerment. Many participants also described a *personal activation* ranging from self-reflection to taking action, i.e. seeking support. The themes summarised in step 2 were often found to be mutually reinforcing, with participants emphasising different themes in varying degrees of intensity. All themes are described in detail in the following sections.

3.2.1. Recognition of depressive symptoms: from denial to awareness

Many participants discussed recognising their depressive symptoms, predominantly as a reaction to the screening questions. By seeing themselves reflected or 'mirrored' by the questions, they perceived their distress or current life problems more intensely. Further, they became aware of symptoms that they did not consciously perceive before. Often, participants reported to have 'ignored' or 'played down' symptoms prior to the screening:

> "There were questions where I wasn't aware before that it bothers me or that it affects me, [... for example] eating behaviour. [...] And sometimes you don't want to be aware of it, you often know it, but you talk yourself out of it. But when you then answer [the questions], then you realise 'oh no - there's something wrong'." (P2, female, 20–29 years, no feedback)

In this context, participants described the questions as 'eye-opening', 'awakening' or as leading to a sense of 'realisation' of the severity of their condition. This included the recognition that they did not feel well, and the classification of their condition as 'not normal' or opposing to how they should ideally feel. Expanding on that, some participants reflected on having incorrectly trivialised or normalised their symptoms before:

> "So, you realised, okay, maybe I've been telling myself all this time 'this is okay', but actually it's not at all." (P26, male, 30–39 years, tailored feedback [not remembered])

Furthermore, completing the screening questions was sometimes described as prompting or 'forcing' a way of introspection that participants would not have come up with on their own:

> 'Because [...] you are sort of [...] stuck with your head in the sand and you don't know where the front and back are. And then there were these very clear, simple questions that no one had asked you before, where you suddenly thought about it in a completely different way.' (P7, female, 30–39 years, no feedback)

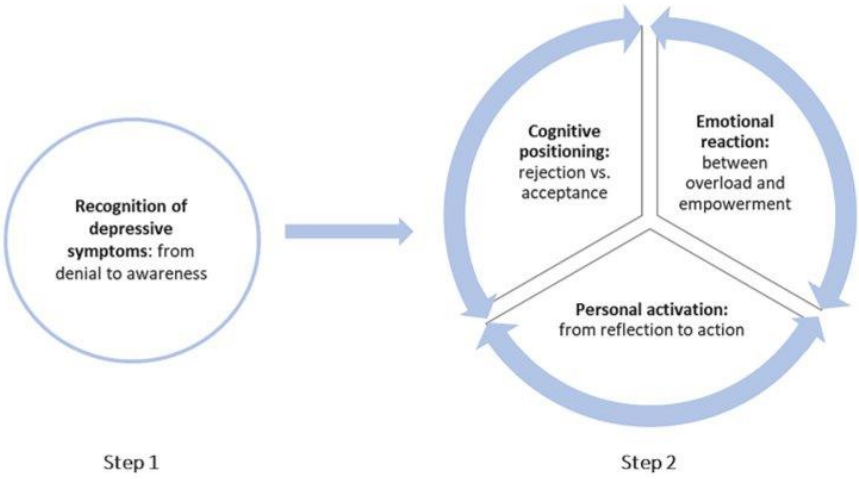


Fig. 1. The experience of online depression screening as a two-step process: Step 1 is the initial reaction to the screening procedure; step 2 describes a subsequent self-explorative process, comprising mutually reinforcing themes.

3.2.2. Cognitive positioning: rejection vs. acceptance

As a response to the screening, many participants reported to have reflected on how to position themselves to a possibly new view of themselves. Most often this referred to the question of whether to reject or accept an illness- or depression-related identity. In that respect, some participants weighed up arguments for and against the possibility of being depressed and remained undecided:

> 'Well, of course I was shocked at first, because it really seemed like depression. [...] But I still knew that it wasn't that bad yet, because I wasn't quite as limited in my everyday life as I would be if I had a severe depression or something like that.' (P10, male, 20–29 years, tailored feedback)

Others rejected an illness-related identity by drawing on an own inner standard, such as a comparison with more severe episodes in the past or the attribution of symptoms on external conditions:

> 'So you already have this feeling that this is not your normal state and that it doesn't sound so good. But I would play it down in my head: Well, everyone is feeling a bit like that [during the pandemic], it's normal at this point and it will automatically go away again.' (P1, female, 20–29 years, standard feedback).

Many of the participants, however, discussed tending to accept a depression-related identity. This acceptance was provoked by both the 'sum of the screening questions' and the explicit feedback. In this context, some participants reflected on a prior cognitive incompatibility between their 'happy' or 'strong' self-image and the possibility of having a depression or 'being in need'. As they reported, the screening facilitated the integration and acceptance of both. Further, some participants described this acceptance as relieving and helpful:

> "So, for me it just became clear 'okay, maybe I am that person now'. Before, it was always 'no, it's definitely not depression, it can't be, I'm always in such a good mood'. But then it was the confirmation 'no, maybe I'm not in a good mood sometimes, maybe I'm just a bit depressed'. [...] And this self-acceptance, this accepting of the, let me express it as an illness, or of the limitations, has already helped me to 'find myself again'." (P10, male, 20–29 years, tailored feedback)

3.2.3. Emotional reaction: between overload and empowerment

The screening process was often reported to elicit diverse and partly intense emotional reactions. At large, participants experienced ambivalent feelings: a first 'shock' or surprise about the realisation of their condition was mostly followed and outweighed by empowering emotions. These included perceptions of 'not being alone' or 'being seen' through the screening questions, and, most often, a feeling of relief. This relief was mainly related to two aspects: First, some participants reported that having an explanation for their condition opened up ideas of how to improve their situation and inspired confidence and hope. Second, relief was often explicated to relate to a validation of the participants' intuition that their condition is 'really an illness' or 'really severe'. In this context, participants discussed that they had questioned this intuition before because they feared to 'imagine' their symptoms or to be to blame for their condition themselves:

> 'Well, there was simply this realisation that I'm not imagining it, that I'm not a hypochondriac, but that it's simply real, and that I can work with it. That has already had a positive influence on me.' (P15, female, 50–65 years, tailored feedback).

Several participants also reported rather challenging emotions. Realising the seriousness of their condition induced for example sadness and self-pity. Further, some participants elaborated on feeling 'uncomfortable' or 'oppressed' as they regretted 'not having seen it come', 'not having cared enough for themselves' or 'not having taken their condition seriously':

> "[…] a bit of an oppressive feeling that maybe I played it down before and didn't really deal with it. [...] So to say I acted against my own feelings. Or that I didn't take it so seriously. [...]. Um, yes, and at that moment I thought 'Well, that was actually not so good, not to address it directly, but to always swallow it down. Um, and to stay in the routines of everyday life'. So to say, this functioning in everyday life." (P11, female, 20–29 years, tailored feedback)

Beyond that, a small number of participants offered accounts of how the screening questions triggered very intense negative emotions and acute distress, that was accompanied by memories from the past. These participants also reported that the regulation of their negative emotions required time-consuming engagement with self-defined coping activities:

> 'So it took quite a while [to get away from the questions]. It wasn't all done straight away with music, but I had to go on the rowing machine for another hour to get myself back on track to some extent.' (P18, male, 50–65 years, standard feedback [not remembered])

3.2.4. Personal activation: from reflection to action

Both screening questions and feedback were described as a trigger or 'catalyst' for a personal activation. This comprised a self-reflective process: participants highlighted that being confronted with the screening questions changed their perspective on themselves and prompted a partly first-time reflection on reasons for their condition. In this context, several participants reported to have come up with psychosocial explanations for their symptoms. As a consequence, they have often questioned their current way of living and dealing with stressors:

> 'The questions lead you to ask specifically where you stand and where you want to go [in life]. [And] no matter what the complaints are, especially if they are psychological, they certainly always have a cause. But under normal daily conditions, you very rarely question such things.' (P12, male, 50–60 years, standard feedback)

Further, participants described engaging in cognitive solution-seeking. Participants receiving only screening questions stated that 'realising' their condition enabled them to identify starting points for change, without however naming these. The focus in participants receiving feedback was more normative and specific: realising that their condition was 'not normal' prompted internal appeals that they 'should' or 'need to' change something, such as seeking a health professional:

> "I think that just showed me 'okay, it's serious' and not something normal and not something you should ignore. And I also thought about maybe going to therapy again." (P13, male, 20–30 years, standard feedback)

Finally, many participants reported to have indeed engaged in active forms of support-seeking. These included self-management activities or self-care (following the screening questions), as well as talking with friends and family or seeking professional help, such as talking to their general practitioner or seeking a psychotherapist (following the feedback):

> "It was definitely good to have [the feedback] in front of my eyes again, somehow, because I think I didn't take my symptoms so seriously at the time, and it was good to see it again in black and white. And I think that was also one of the reasons why I went to therapy again, because then I realised 'okay, it's not normal after all'." (P13, male, 20–29 years, standard feedback)

4. Discussion

Despite growing public use and promotion of online depression screening, current evidence on effects and mechanisms is inconclusive and omits the perspectives of those affected. In this qualitative follow-up study of an RCT on feedback after online depression screening, we aimed to further a comprehensive understanding of online depression screening by exploring how adults with undiagnosed but suspected depressive disorder experience the screening process.

Our results suggest that online depression screening is experienced as a complex two-step process: As a first step, screening prompted the recognition of depressive symptoms by reducing denial and enhancing awareness of symptoms. As a second step, most participants engaged in a self-explorative process encompassing up to three themes: a cognitive positioning towards a potential illness-identity, emotional reactions between empowerment and overload, and/or a personal activation ranging from self-reflection to action. Importantly, participants did not experience all of the described phenomena in the same way. Rather, the focus, the intensity and the perceived valence of the experiences varied across participants. For most, the screening was experienced in a positive way: it enhanced validation and self-understanding, helped to integrate and accept an illness-related self-identity, and/or enabled solution- and support-seeking. For a minority of participants, on the other hand, the screening process elicited negative emotions and acute distress that was challenging to cope with. Lastly, it should be noted that recognition of symptoms and subsequent self-exploration were reported both by participants who received feedback on their condition and by those who answered only the screening questions.

The findings on both benefits and negative effects of the screening process are consistent with prior qualitative research. In the above mentioned study on online depression screening in young adults, participants also reported ambivalent emotional reactions such as validation and shock, as well as actions to manage symptoms such as seeking support (Kruzan et al., 2022). In studies on paper-pencil- instead of internet-based depression screening in primary care or postnatal settings, participants further highlighted an increased awareness of symptoms and a deeper self-understanding. However, screening was also perceived as a personal intrusion, induced a conflict with the self-image, and elicited a rejection of the 'diagnosis' (Dowrick et al., 2009; Shakespeare et al., 2003; Wittkampf et al., 2008).

Another issue named by participants relates to the denial or normalisation of symptoms. Participants described that by forcing the recognition of symptoms, the screening process helped them to overcome normalisation. This links to findings of a qualitative synthesis which showed that delay before help-seeking in depression is often due to normalisation, denial or avoidance of symptoms (Doblyte and Jimenez-Mejias, 2017). Taken together, these findings suggest that recognising the severity of the condition, as opposed to normalising it, appears to be necessary to induce sufficient motivation for change. As such, the recognition of symptoms might be a crucial mechanism of change in (online) depression screening.

In extension to prior research focusing on the mere description of individuals' experiences, we conceptualised participants' experiences as a process leading towards some form of 'activation'. This understanding may be theoretically corroborated by existing behavioural theories such as the Transtheoretical Model of Behaviour Change (TTM; Prochaska et al., 2015). The TTM characterises behaviour change as a series of stages that at large resemble the steps symptom recognition and self-exploration (including the theme personal activation) found in this study: precontemplation (no awareness of need for change), contemplation (some awareness of need for change), preparation, action (taking steps towards change), and maintenance. The TTM further assumes that individuals can enter at any stage and often progress through stages in a nonlinear manner. These assumptions are likewise compatible with our findings and could provide an approach to explain differences in experiences of screening questions and/or feedback between participants.

4.1. Practical and research implications

Traditionally, (online) depression screening was mainly conceptualised as a pathway towards help-seeking, therefore as 'a means to another end'. However, the current study suggests that for many individuals the screening process has a direct subjective benefit itself. This includes meeting individuals' emotional needs for validation and empowerment, self-reflection and self-understanding, and an adaptive positioning towards an illness-related self-identity. To examine generalisability, there is a need for future quantitative research based on patient-oriented outcome measures, ideally assessed directly after the screening. Furthermore, as participants in this study reported to have benefitted from the screening in different ways, future studies should examine how to match the screening process to the different individuals' needs. In this context, the outlined TTM might be a helpful framework to tailor screening interventions specifically to the stage at which an individual enters the screening process.

Altogether, the current findings regarding subjective benefits of online depression screening can help explain the recent public demand for it. Further, they may inform an early and economic provision of low-threshold support for individuals with undiagnosed but suspected depressive disorder.

Of note, benefits of online depression screening were only reported by a sub-sample of participants. On the contrary, the screening procedure also prompted negative emotions and acute distress, which may be categorised as negative effects. Indeed, the risk of negative effects in (online) depression screening is increasingly discussed (Duckworth and Gilbody, 2017; Ryan and Wilson, 2008; Thombs et al., 2012), but research on this subject remains missing (O'Connor et al., 2023). Qualitative findings related to an online intervention for treating depression, however, show similar results: participants described psychological and physical feelings of discomfort attributed to gaining awareness of their condition, to facing negative memories, or to a perceived lack of (therapist) support (Fenski et al., 2021). These and our findings point to the relevance of better understanding the prevalence and the clinical significance of negative effects in online depression screening and, most importantly, of focusing on how these negative consequences can be mitigated for those affected. This is of particular importance as online depression screening is already widely available.

4.2. Limitations

The results of this study should be interpreted in light of the following limitations. First, the study was announced as a survey on stress and well-being and not explicitly called for those seeking online depression screening. Further, both the RCT sample and the interview subsample were self-selected and educated above average. Thus, individuals who participated in the RCT may differ from those using public online depression tests, and participants interested in this follow-up may have been more positive about the screening process or vice versa. However, as the aim of this study was not representativeness, as maximum variation sampling regarding the pre-defined criteria was achieved, and as data analysis resulted in contradictory perspectives, we consider the collected data sufficient. Second, the interview took place approximately six months after screening. Although initial interview questions aimed at helping to recall the screening situation, it cannot be ruled out that participants' memories of the screening process may have been biased. Third, beyond a substantial overlap of experiences across study arms, findings might also indicate differences in the weighting of the reported themes - with the screening questions tending to be associated more with symptom recognition and the feedback of results more with self-exploration. Unfortunately, the design of this study does not allow for drawing valid conclusions on differential effects. For the same reason, possible relationships between mentioned themes and a previous depression diagnosis, that was present in almost half of the participants, could not be examined. Fourth, the sample was recruited from an

RCT. In contrast to public screening practice, participants were paid more attention by repeated surveys and interviews on their mental health, which might have biased their memory of the screening process. Further, the RCT was conducted partly during the COVID-19 pandemic. Although only a minority of participants elaborated on this in their interviews, this might have influenced the participants' clinical characteristics as well as reported experiences. Lastly, online depression screening and feedback were provided in a particular format, so results might not generalise across other public depression screening. It will be important for future work to examine if the pattern of experiencing online depression screening found in this study can be corroborated in naturalistic settings.

4.3. Conclusion

This study furthers a comprehensive understanding of online depression screening. It outlines that screening with and without feedback of results can be experienced as complex two-step process promoting the recognition of depressive symptoms and a subsequent self-exploration. While few participants reported negative effects, the majority described the screening process as insightful, empowering, and activating. Further research should determine to what extent online depression screening may be used as a standalone form of low-threshold support for individuals with undiagnosed depressive disorders, while focusing as well more on potential negative effects.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.invent.2023.100685.

Funding

This work was funded by the German Research Foundation as part of the underlying DISCOVER RCT (grant number: 424162019). We further acknowledge financial support from the Open Access Publication Fund of UKE - Universitätsklinikum Hamburg-Eppendorf and DFG – German Research Foundation.

Compliance with ethical standards

The underlying DISCOVER RCT was approved by the Ethics Committee of the Hamburg Medical Association in July 2019 (reference: PV7039); the present trial was approved by the Ethics Committee of the University Medical Center in June 2021 (reference: 0337).

CRediT authorship contribution statement

SK and BL obtained funding for the underlying RCT (DISCOVER). FS and SK developed the study concept and were engaged in data analysis (re-coding, theme development and theme naming). FS led data collection and analysis and wrote the first draft of the manuscript. SK and BL critically revised the draft for important intellectual content. All authors gave approval of the version published.

Declaration of competing interest

The authors declare that they have no competing interests.

Acknowledgements

We would like to express our sincere gratitude to all participants who generously offered their time and experiences. We would further like to thank Celia Kumpmann who contributed to the organisation and recruitment of the study, as well as Celia Kumpmann, Manon Even, Carlotta Reese, Tarik Korkutan, Simon Kelch, Antonia Thiele, and Hannah Jahn for transcribing the interviews. We are also grateful for Amina Kuby and Albert Anoschin for their assistance with proofreading.

References

Batterham, P.J., Calear, A.L., Sunderland, M., Carragher, N., Brewer, J.L., 2016. Online screening and feedback to increase help-seeking for mental health problems: population-based randomised controlled trial. BJ Psych, Open 2 (1), 67–73. https://doi.org/10.1192/bjpo.bp.115.001552.

Berger, M., Wagner, T.H., Baker, L.C., 2005. Internet use and stigmatized illness. Soc. Sci. Med. 61 (8), 1821–1827. https://doi.org/10.1016/j.socscimed.2005.03.025.

BinDhim, N.F., Alanazi, E.M., Aljadhey, H., Basyouni, M.H., Kowalski, S.R., Pont, L.G., Alhawassi, T.M., 2016. Does a mobile phone depression-screening app motivate mobile phone users with high depressive symptoms to seek a health care professional's help? J. Med. Internet Res. 18 (6), e156 https://doi.org/10.2196/jmir.5726.

Braun, V., Clarke, V., 2006. Using thematic analysis in psychology. Qual. Res. Psychol. 3 (2), 77–101. https://doi.org/10.1191/1478088706qp063oa.

Braun, V., Clarke, V., 2019. Reflecting on reflexive thematic analysis. Qual. Res. Sport, Exerc. Health 11 (4), 589–597. https://doi.org/10.1080/2159676X.2019.1628806.

Braun, V., Clarke, V., 2021. One size fits all? What counts as quality practice in (reflexive) thematic analysis? Qual. Res. Psychol. 18 (3), 328–352. https://doi.org/10.1080/14780887.2020.1769238.

Doblyte, S., Jimenez-Mejias, E., 2017. Understanding help-seeking behavior in depression: a qualitative synthesis of patients' experiences. Qual. Health Res. 27 (1), 100–113. https://doi.org/10.1177/1049732316681282.

Dowrick, C., Leydon, G.M., McBride, A., Howe, A., Burgess, H., Clarke, P., Kendrick, T., 2009. Patients' and doctors' views on depression severity questionnaires incentivised in UK quality and outcomes framework: qualitative study. BMJ 338, b663. https://doi.org/10.1136/bmj.b663.

Dresing, T., Pehl, T., 2015. Manual (on) Transcription: transcription Conventions, Software Guides and Practical Hints for Qualitative Researchers: dr. dresing et pehl GmbH.

Du, N., Yu, K., Ye, Y., Chen, S., 2017. Validity study of patient health Questionnaire-9 items for internet screening in depression among Chinese university students. Asia Pac. Psychiatry 9 (3). https://doi.org/10.1111/appy.12266.

Duckworth, K., Gilbody, S., 2017. Should Google offer an online screening test for depression? BMJ 358, j4144. https://doi.org/10.1136/bmj.j4144.

Eichenberg, C., Wolters, C., Brähler, E., 2013. The internet as a mental health advisor in Germany—results of a national survey. PloS One 8 (11), e79206. https://doi.org/10.1371/journal.pone.0079206.

Erbe, D., Eichert, H.C., Rietz, C., Ebert, D., 2016. Interformat reliability of the patient health questionnaire: validation of the computerized version of the PHQ-9. Internet Interv. 5, 1–4. https://doi.org/10.1016/j.invent.2016.06.006.

Fenski, F., Rozental, A., Heinrich, M., Knaevelsrud, C., Zagorscak, P., Boettcher, J., 2021. Negative effects in internet-based interventions for depression: a qualitative content analysis. Internet Interv. 26, 100469 https://doi.org/10.1016/j.invent.2021.100469.

GBD 2019 Mental Disorders Collaborators, 2022. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet Psychiatry 9 (2), 137–150. https://doi.org/10.1016/s2215-0366(21)00395-3.

Hassem, T., 2022. Evaluating the efficacy of an online depression screening tool in South Africa: a pilot study. S. Afr. J. Psychiatry 28, 1687. https://doi.org/10.4102/sajpsychiatry.v28i0.1687.

Jacobson, N.C., Yom-Tov, E., Lekkas, D., Heinz, M., Liu, L., Barr, P.J., 2022. Impact of online mental health screening tools on help-seeking, care receipt, and suicidal ideation and suicidal intent: evidence from internet search behavior in a large U.S. cohort. J. Psychiatr. Res. 145, 276–283. https://doi.org/10.1016/j.jpsychires.2020.11.010.

Kohlmann, S., Lehmann, M., Eisele, M., Braunschneider, L.E., Marx, G., Zapf, A., Lowe, B., 2020. Depression screening using patient-targeted feedback in general practices: study protocol of the German multicentre GET.FEEDBACK.GP randomised controlled trial. BMJ Open 10 (9), e035973. https://doi.org/10.1136/bmjopen-2019-035973.

Kraus, C., Kadriu, B., Lanzenberger, R., Zarate Jr., C.A., Kasper, S., 2019. Prognosis and improved outcomes in major depression: a review. Transl. Psychiatry 9 (1), 127.

Kroenke, K., Spitzer, R.L., Williams, J.B.W., 2001. The PHQ-9. J. Ge. Intern. Med. 16 (9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x.

Kruzan, K.P., Meyerhoff, J., Nguyen, T., Mohr, D.C., Reddy, M., Kornfield, R., 2022. "I wanted to see how bad it was": online self-screening as a critical transition point among young adults with common mental health conditions. In: Proc SIGCHI Conf Hum Factor Comput Syst, 2022. https://doi.org/10.1145/3491102.3501976.

Levitt, H.M., Bamberg, M., Creswell, J.W., Frost, D.M., Josselson, R., Suarez-Orozco, C., 2018. Journal article reporting standards for qualitative primary, qualitative meta-analytic, and mixed methods research in psychology: the APA publications and communications board task force report. Am. Psychol. 73 (1), 26–46. https://doi.org/10.1037/amp0000151.

Löwe, B., Blankenberg, S., Wegscheider, K., König, H.-H., Walter, D., Murray, A.M., Kohlmann, S., 2016. Depression screening with patient-targeted feedback in cardiology: DEPSCREEN-INFO randomised clinical trial. Brit. J. Psychiat 1–8.

Löwe, B., Spitzer, R.L., Zipfel, S., Herzog, W., 2002. Gesundheitsfragebogen für Patienten (PHQ-D). Manual und Testunterlagen. Universitätsklinikum Heidelberg.

Martin-Key, N.A., Spadaro, B., Funnell, E., Barker, E.J., Schei, T.S., Tomasik, J., Bahn, S., 2022. The current state and validity of digital assessment tools for psychiatry: systematic review. JMIR Ment. Health 9 (3), e32824. https://doi.org/10.2196/32824.

Miller, P., Newby, D., Walkom, E., Schneider, J., Li, S.C., Evans, T.-J., 2021. The performance and accuracy of depression screening tools capable of self-

administration in primary care: a systematic review and meta-analysis. Eur. J. Psychiatry 35 (1), 1–18. https://doi.org/10.1016/j.ejpsy.2020.10.002.

Mitchell, A.J., Vaze, A., Rao, S., 2009. Clinical diagnosis of depression in primary care: a meta-analysis. Lancet 374 (9690), 609–619.

O'Connor, E.A., Perdue, L.A., Coppola, E.L., Henninger, M.L., Thomas, R.G., Gaynes, B. N., 2023. Depression and suicide risk screening: updated evidence report and systematic review for the US preventive services task force. JAMA 2023 329 (23), 2068–2085.

Palinkas, L.A., Horwitz, S.M., Green, C.A., Wisdom, J.P., Duan, N., Hoagwood, K., 2015. Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. Adm. Policy Ment. Health Ment. Health Serv. Res. 42 (5), 533–544. https://doi.org/10.1007/s10488-013-0528-y.

Prochaska, James O., R., C.A., Evers, Kerry E., 2015. The transtheoretical model and stages of change. In: Health Behavior: Theory, Research, and Practice, 97 (2015).

Ryan, A., Wilson, S., 2008. Internet healthcare: do self-diagnosis sites do more harm than good? Expert Opin. Drug Saf. 7 (3), 227–229. https://doi.org/10.1517/14740338.7.3.227.

Shakespeare, J., Blake, F., Garcia, J., 2003. A qualitative study of the acceptability of routine screening of postnatal women using the Edinburgh Postnatal Depression Scale. Br. J. Gen. Pract. 53 (493), 614–619.

Sikorski, F., König, H.-H., Wegscheider, K., Zapf, A., Löwe, B., Kohlmann, S., 2021. The efficacy of automated feedback after internet-based depression screening: study protocol of the German, three-armed, randomised controlled trial DISCOVER. Internet Interv. 25, 100435 https://doi.org/10.1016/j.invent.2021.100435.

Thombs, B.D., Coyne, J.C., Cuijpers, P., de Jonge, P., Gilbody, S., Ioannidis, J.P., Ziegelstein, R.C., 2012. Rethinking recommendations for screening for depression in primary care. Cmaj 184 (4), 413–418. https://doi.org/10.1503/cmaj.111035.

Tong, A., Sainsbury, P., Craig, J., 2007. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. International J. Qual. Health Care 19 (6), 349–357. https://doi.org/10.1093/intqhc/mzm042.

Trautmann, S., Beesdo-Baum, K., 2017. The treatment of depression in primary care: a cross-sectional epidemiological study. Dtsch. Arztebl. Int. 114 (43), 721. https://doi.org/10.3238/arztebl.2017.0721.

UNESCO Institute for Statistics, 2012. International Standard Classification of Education: ISCED 2011, UIS, Montreal. https://doi.org/10.15220/978-92-9189-123-8-en.

Wang, P.S., Berglund, P., Olfson, M., Pincus, H.A., Wells, K.B., Kessler, R.C., 2005. Failure and delay in initial treatment contact after first onset of mental disorders in the National Comorbidity Survey Replication. Arch. Gen. Psychiatry 62 (6), 603–613. https://doi.org/10.1001/archpsyc.62.6.603.

Wittkampf, K.A., van Zwieten, M., Smits, F.T., Schene, A.H., Huyser, J., van Weert, H.C., 2008. Patients' view on screening for depression in general practice. Fam. Pract. 25 (6), 438–444. https://doi.org/10.1093/fampra/cmn057.

7.4.1. Publication IV: Supplementary Material

Supplemetary Table 1

Interview questions.

1. Why did you decide to participate in the DISCOVER study six months ago?
2. How would you describe the symptoms you reported in the first DISCOVER survey six months ago?
3. If you put yourself back into the situation when you were filling in the survey/receiving the feedback, how did you experience that in that very moment (probes on emotions, cognitions, behaviours and time course)?
4. How did your perception or experience of your symptoms change since then?
5. Which impact, if any, did the study/feedback have on your life (if feasible: influence on engagement with symptoms)?
6. How did you, in general, deal with your symptoms in the last six months (if feasible: helpfulness of reported strategies)?
7. What is your general opinion on getting feedback after answering such questions on mental health?
8. When thinking of the feedback, what do you remember most?
9. Regarding the feedback, what did you find helpful/unhelpful – and why?
10. What would you improve if you could design the feedback yourself?
11. What is your spontaneous opinion of online depression tests?
12. What do you think might be advantages/disadvantages of online depression tests?
13. How trustworthy do you evaluate online depression tests?
14. Is there anything else you would like to add?

Supplementary Table 2

COnsolidated criteria for REporting Qualitative research (COREQ): 32-item checklist.

No. Item	Guide questions/description	Reported
Domain 1: Research team and reflexivity		
Personal Characteristics		
1. Interviewer/ facilitator	Which people conducted the interview or focus group?	Franziska Sikorski Sebastian Kohlmann
2. Credentials	What were the researcher's credentials? E.g. PhD, MD	FS: M.Sc. SK: PhD
3. Occupation	What was their occupation at the time of the study?	FS: junior researcher SK: senior researcher
4. Gender	Was the researcher male or female?	FS: female SK: male
5. Experience and training	What experience or training did the researcher have?	FS: training, exchange and supervision by qualitative researchers SK: training, experience in qualitative research
Relationship with participants		
6. Relationship established	Was a relationship established prior to study commencement?	Four participants had a short clinical telephone interview with FS approximately six months before the qualitative interview. FS did not know that when conducting the qualitative interviews.
7. Participant knowledge of the interviewer	What did the participants know about the researcher? e.g. personal goals, reasons for doing the research	Reasons for doing the interviews were known and stated at beginning of the interviews.
8. Interviewer characteristics	What characteristics were reported about the interviewer/facilitator? e.g. Bias, assumptions, reasons and interests in the research topic	Participants were told that the interviewer is interested in their (positive and negative) experiences of the overarching study for her PhD project.
Domain 2: study design		
Theoretical framework		
9. Methodological orientation and Theory	What methodological orientation was stated to underpin the study? e.g. grounded theory, discourse analysis, ethnography, phenomenology, content analysis	Reflexive thematic analysis within an critical realist theoretical framework; experiential orientation; inductive and both semantic and latent coding
Participant selection		
10. Sampling	How were participants selected? e.g. purposive, convenience, consecutive, snowball	Purposeful sampling (maximum variation)
11. Method of approach	How were participants approached? e.g. face-to-face, telephone, mail, email	Telephone (follow-up interview of underlying RCT), followed by email
12. Sample size	How many participants were in the	$N = 26$

	study?	
13. Non-participation	How many people refused to participate or dropped out? Reasons?	Of those consenting to participate, none refused or dropped out.
Setting		
14. Setting of data collection	Where was the data collected? e.g. home, clinic, workplace	At home or at the workplace.
15. Presence of non-participants	Was anyone else present besides the participants and researchers?	No.
16. Description of sample	What are the important characteristics of the sample? e.g. demographic data, date	Participants formerly participated in an online RCT on the efficacy of depression screening; selected characteristics are reported (see Supplementary Table 3).
Data collection		
17. Interview guide	Were questions, prompts, guides provided by the authors? Was it pilot tested?	Questions and prompts are provided (Supplementary Table 1). Pilot testing was conducted within the research team and with one study participant, resulting in small modifications.
18. Repeat interviews	Were repeat interviews carried out? If yes, how many?	No.
19. Audio/visual recording	Did the research use audio or visual recording to collect the data?	Audio recording was used.
20. Field notes	Were field notes made during and/or after the interview or focus group?	Field notes were made if special features have attracted attention.
21. Duration	What was the duration of the interviews or focus group?	M = 37:47 minutes, range: 15:20 minutes to 1:13:56 hours
22. Data saturation	Was data saturation discussed?	Sample size determination (maximum variation sampling) was discussed (as reflexive thematic analysis does not require data saturation; see Braun & Clarke, 2021).
23. Transcripts returned	Were transcripts returned to participants for comment and/or correction?	No.
Domain 3: analysis and findings		
Data analysis		
24. Number of data coders	How many data coders coded the data?	1 (in regular supervision)
25. Description of the coding tree	Did authors provide a description of the coding tree?	No.
26. Derivation of themes	Were themes identified in advance or derived from the data?	Themes were derived inductively from the data.
27. Software	What software, if applicable, was used to manage the data?	MAXQDA software (2022)

28. Participant checking	Did participants provide feedback on the findings?	No.
Reporting		
29. Quotations presented	Were participant quotations presented to illustrate the themes/findings? Was each quotation identified (e.g. ID)?	Yes, quotations including the ID of the participants are reported.
30. Data and findings consistent	Was there consistency between the data presented and the findings?	Yes.
31. Clarity of major themes	Were major themes clearly presented in the findings?	Yes, all themes relating to the research question are presented.
32. Clarity of minor themes	Is there a description of diverse cases/discussion minor themes?	Yes, diverse cases are discussed.

Note. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. International Journal for Quality in Health Care. 2007. Volume 19, Number 6: pp. 349 – 357.

Supplementary Table 3

Selected characteristics per participant.

Participant	Gender	Age range	Allocated study arm (remembered by the participant)	Depression severity (PHQ-9 score)	Depression diagnosis in the past
1	female	20 - 29	standard feedback (no)	moderate (10)	no
2	female	20 - 29	no feedback	severe (17)	no
3	male	40 - 49	no feedback	severe (17)	yes
4	female	20 - 29	no feedback	severe (20)	yes
5	female	20 - 29	standard feedback (yes)	moderate (12)	no
6	female	20 - 29	Standard feedback (yes)	severe (16)	yes
7	female	30 - 39	no feedback	severe (15)	no
8	female	20 - 29	Standard feedback (yes)	moderate (13)	yes
9	female	40 - 49	tailored feedback (yes)	moderate (13)	yes
10	male	20 - 29	tailored feedback (yes)	moderate (13)	no
11	female	20 - 29	tailored feedback (yes)	severe (19)	no
12	male	50 - 65	standard feedback (yes)	moderate (10)	no
13	female	20 - 29	standard feedback (yes)	moderate (10)	yes
14	female	30 - 39	tailored feedback (yes)	severe (17)	yes
15	female	50 - 65	tailored feedback (yes)	severe (26)	yes
16	male	40 - 49	tailored feedback (no)	moderate (11)	yes
17	female	50 - 65	standard feedback (no)	moderate (12)	no
18	male	50 - 65	standard feedback (no)	moderate (14)	yes
19	male	30 - 39	no feedback	moderate (10)	no
20	male	50 - 65	no feedback	severe (24)	yes
21	male	30 - 39	standard feedback (yes)	severe (17)	no
22	female	50 - 65	no feedback	severe (21)	yes
23	male	50 - 65	tailored feedback (no)	severe (20)	no
24	female	50 - 65	standard feedback (yes)	moderate (11)	no
25	male	30 - 39	standard feedback (no)	severe (23)	no
26	male	30 - 39	tailored feedback (no)	moderate (10)	no

Note. PHQ-9 = Patient-Health-Questionnaire-9; moderate = 10 ≥ PHQ-9 < 15; severe = PHQ-9 ≥ 15.

8 AUTHOR CONTRIBUTION STATEMENT

Financial support for the DISCOVER project was given by the German Research Foundation (Deutsche Forschungsgemeinschaft) and procured by Prof Dr Sebastian Kohlmann and Prof Dr Bernd Löwe, who also supervised all studies/publications. In all four publications, Franziska Sikorski was involved in project development and organisation, in recruitment, data collection, data cleansing, and data analysis, and contributed to preparation of the manuscripts by writing, reviewing, and editing.

Publication I

Sikorski, F., König, H. H., Wegscheider, K., Zapf, A., Löwe, B., & Kohlmann, S. (2021). The efficacy of automated feedback after internet-based depression screening: Study protocol of the German, three-armed, randomised controlled trial DISCOVER. *Internet Interventions*, 25, 100435.

Franziska Sikorski developed the intervention together with Sebastian Kohlmann, contributed to outcome selection and timing, contributed to study website development, created all figures and tables, wrote the original draft of the manuscript, and revised the manuscript after peer-review**.**

Publication II

Kohlmann, S., Sikorski, F., König, H.-H., Schütt, M., Zapf, A., Löwe, B., (2024). The efficacy of automated feedback after internet-based depression screening (DISCOVER): an observer-masked, three-armed, randomised controlled trial in Germany. *The Lancet Digital Health*, 6(7), e446-e457.

Franziska Sikorski was involved in project organisation, collected the data together with Sebastian Kohlmann, was involved in data curation, cleaned the data, contributed to data analysis, contributed to interpretation of data for the publication, created all figures and tables, critically reviewed and edited the original and revised draft.

Publication III

Sikorski, F., Löwe, B., Daubmann, A., & Kohlmann, S. (under review). Does feedback after online depression screening cause harm? A secondary analysis of negative effects in the randomised controlled DISCOVER trial. *Journal for Medical Internet Research.*

Franziska Sikorski designed the study together with Sebastian Kohlmann and Anne Daubman, collected the data together with Sebastian Kohlmann, cleaned the data, performed the data analyses, created all figures and tables, and wrote the original draft of the manuscript.

Publication IV

Publication IV: Sikorski, F., Löwe, B., & Kohlmann, S. (2023). How adults with suspected depressive disorder experience online depression screening: A qualitative interview study. *Internet Interventions,* 34.

Franziska Sikorski designed the study design together with Sebastian Kohlmann, led the data collection and data analysis, created all figures and tables, wrote the original draft of the manuscript, and revised the draft after peer-review.

DANKSAGUNG

Mein besonderer Dank gilt Herrn Prof. Dr. Sebastian Kohlmann für die Betreuung dieser Arbeit und die jahrelange Begleitung und Unterstützung auf allen Ebenen. Danke für deine Akzeptanz von Ambivalenzen, für deine Zeit, Geduld und Empathie sowie deinen Ansporn in jeweils den richtigen Momenten, dein Vertrauen in mich, deine Gelassenheit sowie deinen klugen, kritischen und doch pragmatischen Blick - es ist kaum aufzählbar, was ich in den letzten Jahren alles von dir lernen konnte.

Ein großer Dank gilt auch Herrn Prof. Dr. Bernd Löwe und Herrn Prof. Dr. Levente Kriston für die Mitbetreuung dieser Arbeit und die wertvollen Anregungen. Herr Löwe, Ihnen darüber hinaus vielen Dank für die Unterstützung und Förderung - die Erfahrungen, die ich in der Klinik für Psychosomatik sammeln durfte, weiß ich sehr zu schätzen.

Ebenfalls bedanken möchte ich mich bei Elisa Rabbiolo, Tarik Korkutan, Ajla Hamidovic und Celia Kumpmann für die unverzichtbare Unterstützung unserer Studie als wissenschaftliche Hilfskräfte und Praktikantinnen sowie bei allen Studienteilnehmer:innen, die unsere Forschung durch ihre Zeit und ihren Mut zur Selbstöffnung erst möglich gemacht haben.

Ein weiterer Dank gilt all meinen Kolleg:innen und Mitdoktorand:innen für die fachliche und auch mentale Unterstützung. Besonders danke ich Lea Braunschneider, für das geteilte Lachen und Weinen im Büro und das gegenseitige Lebenscoaching sowie Natalie Uhlenbusch, für deine Hilfsbereitschaft, deine Herzlichkeit und deine Leidenschaft für die Forschung - du warst und bist mir ein Vorbild.

Danke auch an alle weiteren Menschen, die mich in den letzten Jahren begleitet haben. Allen voran Dalia Luna Linden, die mir ihr Vertrauen in mich geliehen und so den Anstoß für meinen Weg in die Wissenschaft gegeben hat, Alexander Sikorski, meinem Joker in allen Lebens-, Forschungs- und Technologiefragen, und Albert Anoschin – für deine Liebe, deinen ansteckenden Größenwahn und deine kritischen Korrekturen (insbesondere die in letzter Sekunde!). Ein großer Dank gebührt auch meinen Eltern, für eure unermüdliche Unterstützung und alle Anstrengungen, die meinen Weg in dieser unbekannten Welt ebnen sollten. Und zu guter Letzt noch ein Dank an Lampi, für die ermutigende Begleitung durch Berg und Tal und das stets rechte Maß an Zauberei.

LEBENSLAUF

Lebenslauf aus datenschutzrechtlichen Gründen nicht enthalten.

EIDESSTATTLICHE VERSICHERUNG

Ich versichere ausdrücklich, dass ich die Arbeit selbständig und ohne fremde Hilfe, insbesondere ohne entgeltliche Hilfe von Vermittlungs- und Beratungsdiensten, verfasst, andere als die von mir angegebenen Quellen und Hilfsmittel nicht benutzt und die aus den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen einzeln nach Ausgabe (Auflage und Jahr des Erscheinens), Band und Seite des benutzten Werkes kenntlich gemacht habe. Das gilt insbesondere auch für alle Informationen aus Internetquellen.
Soweit beim Verfassen der Dissertation KI-basierte Tools („Chatbots“) verwendet wurden, versichere ich ausdrücklich, den daraus generierten Anteil deutlich kenntlich gemacht zu haben. Die „Stellungnahme des Präsidiums der Deutschen Forschungsgemeinschaft (DFG) zum Einfluss generativer Modelle für die Text- und Bilderstellung auf die Wissenschaften und das Förderhandeln der DFG“ aus September 2023 wurde dabei beachtet.
Ferner versichere ich, dass ich die Dissertation bisher nicht einem Fachvertreter an einer anderen Hochschule zur Überprüfung vorgelegt oder mich anderweitig um Zulassung zur Promotion beworben habe.
Ich erkläre mich damit einverstanden, dass meine Dissertation vom Dekanat der Medizinischen Fakultät mit einer gängigen Software zur Erkennung von Plagiaten überprüft werden kann.

Datum Unterschrift